University of Alberta

Vision techniques applied to pattern extraction and matching for the purpose of 3D back reconstruction

by

Jaishankar Thayyoor

A thesis submitted to the faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

Department of Electrical and Computer Engineering

Edmonton, Alberta

Fall 1998

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-34427-4

Canada

University of Alberta

Library Release Form

Name of Author:	Jaishankar Thayyoor
Title of Thesis:	Vision Techniques applied to pattern extraction and matching for the purpose of 3-D back reconstruction
Degree:	Master of Science
Year this Degree Granted:	1998

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

T. V. Jaishankar

Jaishankar Thayyoor
16,Lady Madhavan Road
Chennai (Madras)
Tamil Nadu 600 034
India.

Date: June. 3, 1998

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled *'Vision techniques applied to pattern extraction and matching for the purpose of 3-D back reconstruction'* submitted by *Jaishankar Thayyoor* in partial fulfillment of the requirements for the degree of *Master of Science*.

N.G.Durdle

V.J.Raso

A.E.Peterson

K.A.Stromsmoe

Date: Jan 31/96

Abstract

A stereo vision system using structured light was developed to automatically acquire the three-dimensional coordinate information of the human trunk surfaces. To improve the pattern extraction and matching of structured light projected on the surface, a three-level coding scheme was developed. The data acquisition process consists of photogrammetric and image-processing techniques such as camera calibration, image rectification, image correlation and correspondence matching. The system was tested using solid models and human subjects. The accuracy and the reliability of the image processing algorithms were investigated and the results show that the system has the potential of becoming a useful clinical tool for the evaluation of scoliosis.

To my parents

Acknowledgements

I would like to express my gratitude to my supervisor, Dr.Nelson G. Durdle, Mr Jim Raso and Prof. A.E. Peterson for their encouragement support and guidance throughout this research. I am obliged to Mr. Doug Hill of the Glenrose Rehabilitation group for his valuable guidance with research and patiently suffering through various drafts of this thesis and providing valuable feedback with respect to both content and style.

I also wish to express my heartfelt thanks to Kees and Shaju for the support they so ever willingly provided. Thanks to Michelle, Carla and Shirley at the Department office for all the help. I would like to acknowledge the help of Lawrence, John, Edmond, Martin, Kevin and Rahman for their enthusiastic support in both technical and non-technical matters.

I also wish to express my gratitude to Dr. (Mrs.) Radha Gourishankar and Dr. V. Gourishankar for the guidance meted at every stage of my stay here. I am obliged to them for all their help- from course work to exam preparation to dropping me back home at the slightest opportunity. Their moral support and encouragement made my stay in Edmonton a memorable one.

I take this opportunity to thank my friends for being what they are and knowing when not to ask, "How's your thesis coming?" Special thanks to my roommates Rohit and Lucio for making my stay in HUB as pleasant as possible. A word of thanks to the Edmonton radio stations CBC and 630CHED for keeping me alert and awake late at night.

I am extremely grateful to my parents and my sister Supritha for their love and encouragement and imparting to me their value of an education. Priyamvatha provided me with invaluable support and inspiration. I owe her more than thanks.

Table of Contents

1 Introduction 1

1.1 Purpose 1

1.2 Motivation 1

1.3 Objectives 2

1.4 Overview 2

2 Background 5

2.1 Scoliosis 5

2.1.1 Surface Topography Measurement 7

2.2 Stereovision and Photogrammetry 8

2.2.1 Camera-projector Geometry 9

2.2.2 Camera Calibration 10

2.2.3 Image Rectification 10

2.2.4 Correspondence Matching 13

2.2.5 Feature Extraction by Template Matching 14

2.3 Literature Survey of Coding Schemes 16

2.3.1 Binary Coding 16

2.3.2 Three-level Coding 21

2.3.3 Colour Coding 23

2.3.4 Summary of Literature Review 25

3 Experiments to identify an appropriate coding scheme 27

3.1 Experimental Setup 27

3.1.1 Calibration Procedure 28

3.1.2 Projector Slide Design 29

3.2 Binary (Two level) Coding Scheme 29

3.3 Three Level Coding 38

3.4 Colour Coding Scheme 43

3.5 Analysis of the Results 44

3.6 Justification for Selecting the Three-level Coding Scheme 45

4 Image Processing and 3D Surface Reconstruction 47

4.1 Image Rectification 47

4.1.1 Validation of the Rectification Procedure 50
4.2 Feature Extraction 53
4.2.1 Template Design 53
4.2.2 Implementation of Feature Extraction by Correlation 57
4.2.3 Validation of Feature Extraction 63
4.3 Correspondence Matching 66
4.3.1 Implementation of the Correspondence Matching 67
4.4 Triangulation and 3D Surface Reconstruction 69
5 System Validation and Testing 72
5.1 Reconstruction of Test Objects 72
5.2 Analysis of Results 85
5.3 Recovery of 3-D Surface Information from the Back Images 86
5.4 Discussion 94
6 Conclusion and Recommendations for future work 98
6.1 Conclusion 98
6.2 Recommendations for Future Work 99
References 102
Appendix 108

List of tables

Table 3.1: Results of camera calibration ---------- 28
Table 4.1: Test data used as input for rectification---------- 51
Table 4.2: Known rectified points of table 4.1---------- 51
Table 4.3: Computed rectified points of table 4.1---------- 52
Table 4.4: Sequence of codes as appears in the captured stereo images ---------- 68
Table 5.1: Results for the curved surface----------76
Table 5.2: Results from the 3 trials conducted---------- 76
Table 5.3: Results for ramp surface---------- 78
Table 5.4: Results obtained for step object---------- 85
Table 5.5: Results for three trials conducted on male subject ---------- 86

List of Figures

Figure 2.1: Patient with Scoliosis ---- 6
Figure 2.2: Cobb angle measurement ---- 6
Figure 2.3: Camera-Projector system setup ---- 9
Figure 2.4: General epipolar geometry ---- 11
Figure 2.5: Rectification of an image ---- 12
Figure 2.6: Arrangement for obtaining the correlation ---- 15
Figure 2.7: When grid is encoded with dots to form PRBA ---- 19
Figure 2.8: Projected binary pattern ---- 20
Figure 2.9: The Illumination pattern is composed of four primitives ---- 20
Figure 2.10: Coding the lighting sections ---- 21
Figure 2.11: Binary line coding using multiple patterns ---- 21
Figure 2.12: Three sinusoidal fringe patterns for phase shift illumination ---- 22
Figure 2.13: The pattern used for the colored lighting system ---- 24
Figure 2.14: The five modified grid primitives ---- 24
Figure 3.1: Step object with horizontal lines ---- 30
Figure 3.2: Shift of lines due to steep changes in depth ---- 30
Figure 3.3: Shows Occlusion ---- 31
Figure 3.4: Vertical and horizontal lines projected on an object ---- 32
Figure 3.5: Projected slide with two pixel wide lines ---- 33
Figure 3.6: Captured image of a 2pixel wide pattern ---- 33
Figure 3.7: Plot of intensity vs number of pixels of projected pattern ---- 34
Figure 3.8: Plot of intensity vs number of pixels on the captured image ---- 34
Figure 3.9: Projected image of 4 pixel wide lines ---- 36
Figure 3.10: Captured image of the 4 pixel wide line ---- 36
Figure 3.11: Plot of 4 pixel wide projected lines ---- 37
Figure 3.12: Plot of intensity vs number of pixels of the captured image ---- 37
Figure 3.13: Three level coding schemes ---- 38
Figure 3.14: A slide created with three level coding ---- 39
Figure 3.15. Captured image of the three level pattern ---- 39
Figure 3.16: Plot of intensity vs number of pixels of the projected pattern ---- 40

Figure 3.17: Plot of intensity vs number of pixels of the captured image---------- 40
Figure 3.18: Detected levels and Projected levels ---------- 41
Figure 3.19: Tracing broken lines with three-level pattern ---------- 43
Figure 3.20: Color coded slide ---------- 44
Figure 3.21: Selected three-level coded pattern ---------- 46
Figure 4.1: The object coordinate system.---------- 48
Figure 4.2: Rectification of two images ---------- 49
Figure 4.3: Tilted image of a plane ---------- 52
Figure 4.4: Rectified image of a plane ---------- 52
Figure 4.5: Captured image of a plane surface ---------- 55
Figure 4.6: Column of captured image ---------- 56
Figure 4.7: Plot of intensity along the column of the captured image---------- 56
Figure 4.8: Correlation results for 54 pixel template ---------- 57
Figure 4.9: Correlation results for an entire column---------- 58
Figure 4.10: Correlation results for 18 pixel template ---------- 60
Figure 4.11: Plot the intensity values of that subcode in the captured image ---------- 62
Figure 4.12: Computer generated test image ---------- 63
Figure 4.13: Results after feature extraction process ---------- 64
Figure 4.14: Computer generated test image ---------- 64
Figure 4.15: Results after features extraction process ---------- 65
Figure 4.16: Stereo Images of a ramp like object ---------- 66
Figure 4.17: Results of the feature extraction process ---------- 66
Figure 4.18: 3-D points of the triangular surface ---------- 70
Figure 4.19: Various steps involved in 3D surface reconstruction ---------- 71
Figure 5.1(a): Top Camera view of the object surface ---------- 73
Figure 5.1(b): Bottom Camera view of the object surface ---------- 73
Figure 5.2(a): Results of feature extraction of the top image ---------- 75
Figure 5.2(b): Results of feature extraction of the bottom image ---------- 75
Figure 5.3: 3D view of the surface from two angles ---------- 77
Figure 5.4: Captured stereo images of the triangular object ---------- 79
Figure 5.5(a): Results of feature extraction from the top image ---------- 80

Figure 5.5(b): Results of feature extraction from the bottom image ---------- 80
Figure 5.6(a): 3-D View of the surface ---------- 81
Figure 5.6(b): 3D view of the ramp surface ---------- 81
Figure 5.7: Stereo images of the step like object ---------- 82
Figure 5.8: Results of feature extraction for step object ---------- 83
Figure 5.9: 3D view of the steps ---------- 84
Figure 5.10: Stereo images of a male subject ---------- 88
Figure 5.11: Results after feature extraction ---------- 89
Figure 5.13: Stereo Images of female subject ---------- 90
Figure 5.14: Results after feature extraction ---------- 90
Figure 5.15: 3D view of the surface ---------- 91
Figure 5.16: Stereo images of 2nd male subject ---------- 92
Figure 5.17: 3D view of the reconstructed surface ---------- 92
Figure 5.18: The top camera image of the 3rd male subject ---------- 93
Figure 5.19: 3-D view of the reconstructed surface ---------- 93
Figure A.1: Flow of the software modules ---------- 110

1 Introduction

1.1 Purpose

This work describes the research and development of a method to improve the pattern extraction and matching of surface topography features and to assist in the evaluation of scoliotic deformities. The aim of this research is to develop an appropriate vision technique that would aid in the development of an automated system that acquires three-dimensional information of the trunk surfaces. This technique involves the use of a structured (coded) light combined with image processing methods for solving the correspondence problem in stereo vision. The objective of this research is to investigate the coding schemes that could be used to improve the automated image capture system.

A fully automated system for recording back surface topography can allow useful back-shape information to be available to the clinician within a very short time, and it provides a viable alternative to the current methods of measuring the deformity associated with scoliosis. The proposed technique is non-contact and poses no risk to the patient.

1.2 Motivation

The Glenrose Rehabilitation Hospital is involved in the assessment of scoliosis. Many three-dimensional data acquisition techniques have been investigated for developing a system to assist clinicians in the evaluation of scoliotic deformities. The back surface is smooth and featureless, so it becomes very difficult to use image processing methods to create a three-dimensional surface model or to identify landmarks for clinical measurements. To overcome this problem, a coded light pattern is projected onto the patient's back using a 35mm slide projector. The features of this pattern are then easier to identify and can be used to define three-dimensional coordinates and to generate surface models. This technique is called: the structured light technique.

Structured light is the most common technique used in machine vision [7], [13], [15], [23], [40], [42]. Numerous coded light techniques have been proposed by researchers, one of which was investigated at the Glenrose Rehabilitation research center

by Liu [27]. This approach uses parallel horizontal lines as a coded pattern and provides three-dimensional coordinates with a resolution of 2mm in the horizontal direction and 13mm in the vertical direction. Improvements in the vertical accuracy are required, since a minimum accuracy of 2mm in all directions is desired for clinical assessment of scoliosis.

1.3 Objectives

The objective of this research is to investigate coding schemes that could be used to improve the automated system described above. This can be achieved by developing a coding pattern for improved detection accuracy and develop intelligent techniques for the pattern extraction and matching of the stereo images. To develop an appropriate coded pattern the various schemes proposed by researchers have to be investigated. Once a suitable coding scheme is chosen, image processing techniques and principles of photogrammetry can be used to solve the correspondence problem in stereo vision.

1.4 Overview

The various coding schemes that have been developed by researchers have specific advantages and disadvantages when applied to surface extraction of human torso information. Most applications involve the reconstruction of objects where edges are well-defined [3], [46]. These methods differ from the method applied to the human back because the human back is predominantly a flat surface and has few distinct features, compared to a face or a cuboid. Another challenge posed by human subjects is that, the stereo camera system has to capture images in a very short time interval to minimize the breathing effects and other movement artifacts of the patient. These and other issues are dealt with in the following chapters.

Chapter 2 provides information on scoliotic deformities and the current clinical methods of diagnosing scoliosis. The need for an automated clinical tool has been investigated and the justification for providing surface topography as a useful technique is presented. Stereo vision and relevant literature in this area are discussed. The photogrammetric techniques involved in the design of the stereo vision system are also

presented. The coding schemes suggested by various researchers are reviewed and the possibility of using some of those techniques is discussed. This chapter also introduces the image processing techniques, such as image rectification and image correlation that will be used to solve the correspondence problem in stereo images.

From the literature survey, some of the promising coding schemes were investigated to identify an appropriate coding scheme, which can be used for the measurement of the surface topography of the (human) back. The coding schemes to be investigated were categorized as binary, three-level and multilevel (or colour) coding schemes. The experiments carried out with the selected coding schemes and the efficiency of each scheme in solving the correspondence problem is presented in Chapter 3. To carry out this investigation, the camera parameters have to be computed by calibration. The calibration procedure and results are also presented in this chapter. The results of the experiments and the justification for selecting the three-level coding scheme for this application (reconstruction of the surface of the back) are provided.

The selected coding scheme adds known features to the surface of the back. The stereo cameras capture the image of the surface of the back along with the deformed coded pattern. Chapter 4 describes the implementation of the image processing techniques that were carried out to extract useful information (features) from the captured stereo images. The various techniques involved from image capture to 3-D surface reconstruction are explained. The software implementation of image rectification, image correlation and the solution to the correspondence-matching problem by feature extraction is presented. The computation of the 3-D coordinates of the matched points by using the technique of triangulation is also described in this chapter.

Chapter 5 deals with the tests conducted with known models to evaluate the performance of the system. The choice of the test models and the results of the tests are presented. The tests conducted on subjects with normal back surface and with mild scoliosis are described and the results obtained from these tests are also provided. The accuracy, resolution, level of automation, repeatability and computational time are discussed. The limitations of the system are identified and the techniques to minimize the errors in measurement are also explained in this chapter.

Chapter 6 presents conclusions and the recommendations for further work.

The system specification and details of the image processing software and are provided in the Appendix.

2 Background

This chapter reviews the background material to this research. Scoliosis and methods of diagnosing this deformity are reviewed. The need for surface topography as an automated clinical tool is justified. An overview of the stereo vision system and its components are provided. The photogrammetric and image processing techniques to be used for this application are introduced. Literature review of structured light coding techniques proposed by researchers is also presented.

2.1 Scoliosis

Scoliosis is the abnormal lateral curvature of the spine with vertebral rotation. Lateral spinal curvature is often accompanied by an asymmetry in the shape of the back (figure 2.1). Rotation of the vertebrae causes the ribcage to distort and a hump is produced on one side of the back. Radiological assessment of the lateral curvature in scoliosis usually includes measurement of the Cobb angle (figure 2.2) which is the angle subtended by the normals to the points of inflexion (maximum curvature) in the spinal midline. 2-4% of the population have scoliosis curves of at least 10 degrees, and 0.1-0.3% of children have curvatures greater than 20 degrees. In adolescents, scoliosis can progress at an alarming rate and have severe effects on the cosmesis of an individual. Some of the common visible characteristics are

- One shoulder may be higher than the other.
- One scapula (shoulder blade) may be higher or more prominent than the other.
- The trunk is shifted over the pelvis.
- One hip may appear to be higher or more prominent than the other is.
- The head is not centered over the pelvis.
- When the patient is examined from behind and asked to bend forward until the spine is horizontal, one side of the back appears higher than the other side.

Figure 2.1: Patient with Scoliosis (*Courtesy: Rehab Tech,. Glenrose Hospital).*

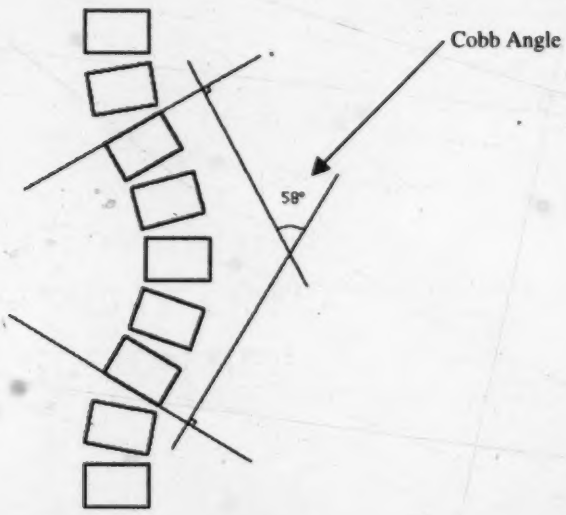


Figure 2.2: Cobb angle measurement

The most common form of scoliosis is *idiopathic* scoliosis, meaning 'spinal curvature without a cause'. In spite of a century of research, little is known about the etiology of idiopathic scoliosis and consequently, there are no preventive measures. Quantification of the surface deformity has enabled investigation of curve progression and effectiveness of surgical treatment. Clinically, the quantification of radiographic images of the curve still relies on the use of Cobb angle to define lateral deformity, however there are three relevant disadvantages to the use of radiography. First it is an invasive technique, second the two-dimensional representation of a structure which is deformed in three-dimensional space is not always adequate; third, it lacks good visualization of the rib cage whose associated deformity is, thus, not well documented.

2.1.1 Surface Topography Measurement

An alternative method for analyzing the extent of scoliosis is to measure the three-dimensional surface deformity of the back. Measurement of surface shape does not usually involve radiological techniques and the harmful effects of ionizing radiation are therefore avoided. Surface topography allows a good description of surface asymmetry, a pathological finding whose detection has great importance in defining the deformed rib cage, giving data related to the consequent respiratory impairment and aesthetic damage. Standard radiography and surface topography can each find its own place in assessing and monitoring scoliosis since each one performs better in areas where the other is weak.

However, at the present time, surface topography cannot replace standard radiography but it can be very useful in the frequent assessment of conservatively managed cosmetic defects. The patient with scoliosis is often most concerned about the perceived visible surface deformity of their back rather than the underlying structural deformity of the spine. Control of the back surface shape may be more important in terms of patient satisfaction than correction of spinal deformity. The time and effort to acquire, process and record clinical data using manual techniques has limited useful application in the scoliosis clinic and in screening programs. A fast accurate and fully integrated system to acquire, process and record surface shape data is therefore desirable.

2.2 Stereovision and Photogrammetry

There are several optical, non-contact approaches to obtain 3D topographical data of objects. Among these are Moiré methods and Structured light techniques using stereo imaging.

Moiré Techniques

A moiré pattern is defined as the low spatial frequency interference pattern created when two gratings with regularly spaced patterns of higher spatial frequency are superimposed on one another. Moiré range-imaging methods are useful for measuring the relative distance to surface points on a smooth surface that does not exhibit depth discontinuities [33]. Depth of field of a moiré system depends on the camera resolution and the object-grating period. The moiré topographic method requires multiple (two) observations to establish the direction of depth change. Therefore longer observation time is required when taking images of a patient.

Structured Light Approach

Triangulation with structured light is a well-established technique for acquiring range pictures of three-dimensional scenes [23], [42]. Structured lighting refers to a technique that involves illuminating the scene with controlled lighting and interpreting the pattern of the projection in terms of the surface geometry of the objects. The use of structured light is aimed at reducing the computational complexity and improving the reliability of the 3-D object recognition process. Other optical techniques such as shape from shading and stereoscopy, gather information from ambient light reflection, while structured lighting methods cast modulated light (light cast with spatially variant patterns) onto the scene to obtain a 'spatial encoding' for analysis. Use of structured light and triangulation is a very popular technique for sensing 3D surface points [5], [6], [10], [14].

From a single image, it is generally not possible to infer unambiguous information about the shape location of 3-D objects in a viewed scene. Passive stereo vision approaches use two (or more) calibrated cameras in distinct locations to reduce ambiguity. Stereo imaging has the advantage of giving more direct, unambiguous and quantitative depth information. Also, it is usable for a wide range of applications besides that of 3D topographical measurement, e.g. applications such as robotic vision [19],

autonomous vehicle control [12], and sight for the blind [11], remote sensing and automated manufacturing [38]. Most approaches to the application of stereo vision utilize the human vision system to establish a model for the camera.

However, with stereo vision systems the difficult problem of determining which points in both images correspond to the same 3-D scene point, or so-called 'correspondence problem' arises. One way to alleviate the correspondence problem is to actively control the scene by projecting known patterns onto the objects. This approach is referred to as 'active stereo vision' or structured light technique.

Liu [27] investigated a structured light approach based on parallel horizontal lines as the projected pattern. This method provides three-dimensional coordinates with a resolution of about 13mm in the vertical direction. The parallel lines used were not densely spaced and this affected the detection accuracy of the features of the back. This system was also limited by occlusion. Further work is required to improve this resolution and thereby improve the performance of the system.

2.2.1 Camera-projector Geometry

The stereo camera system used by Liu consisting of two cameras and a 35mm projector are arranged as shown in figure 2.3.

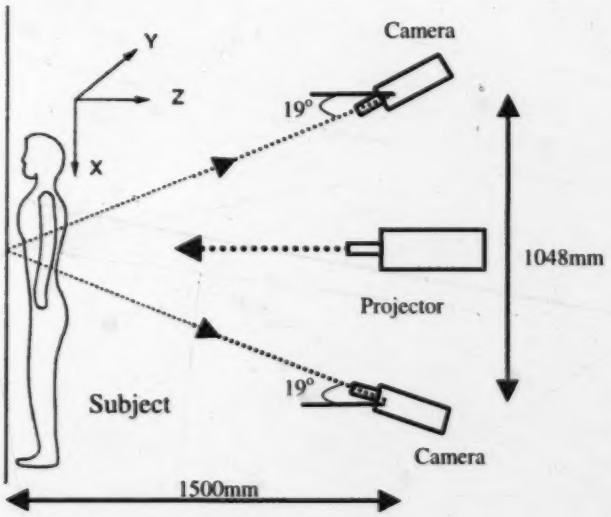


Figure 2.3: Camera-Projector system setup

A computer with image capture facility is used to obtain the stereo images. The projector projected either spatially labeled patterns or colour encoded patterns onto the subject/object. Photogrammetric concepts were used to derive optimal stereo-camera geometry for the system and to establish the calibration parameters of the cameras. The cameras are mounted symmetrically above and below the projector. The base line distance between the two cameras is 1000mm and the horizontal distance between the cameras and subject is about 1400mm-1500mm. Convergence angles of approximately 20^{o} are set for both video cameras. It was found that with these convergence angles and distances, both cameras have the same field of view in the object space and are able to capture a full human trunk surface [27]. Apart from its influence on the range resolution, the angle between the projector and the cameras also influences the amount of shadow observed in the scene.

2.2.2 Camera Calibration

To perform accurate range estimation, the measurement system will first have to be calibrated. During calibration, parameters like position, orientation and focal distances of the cameras have to be estimated. The calibration program (v2stereo.f) developed by A.Peterson solves for thirteen parameters of both cameras in the system [27]. The thirteen parameters are world X, Y, Z coordinates of the camera's perspective center, orientation of the camera (rotation angles ω,ϕ, and κ with respect to the X, Y, Z, axes), focal length, two radial distortion parameters, K1, K2, and two decentering distortion parameters, P1, P2 as well as the principal point coordinates (xp, yp). The calibration procedure is discussed in Chapter 3.

2.2.3 Image Rectification

Most algorithms in computer vision and digital photogrammetry are based on the assumption that digital stereo pair is registered in epipolar geometry [8]. That is, the scan lines of stereo pairs are epipolar lines. This condition is satisfied when the two camera axes of a stereo vision system are parallel to each other and perpendicular to the camera base. In stereo imaging, one can greatly limit the computations required in determining

correspondence of the two images, by constraining the geometry of the cameras during image acquisition. If two balanced, equal focal length cameras are arranged with their axes parallel, then they share a single, common image plane. Any point in the scene will project to two points in that joint image plane, the connection of which produces a line parallel to the baseline between the two cameras. This line is called as an epipolar line (figure 2.4). If the baseline between the two cameras is parallel to an axis of the cameras, then the correspondence needs to be searched only along lines parallel to that axis in the image. However, such an arrangement is difficult to achieve in practice. In this research work the cameras have a converging optical axis (figure 2.3) so that sufficient field of view is obtained. It is therefore necessary to rectify the images in order to process them.

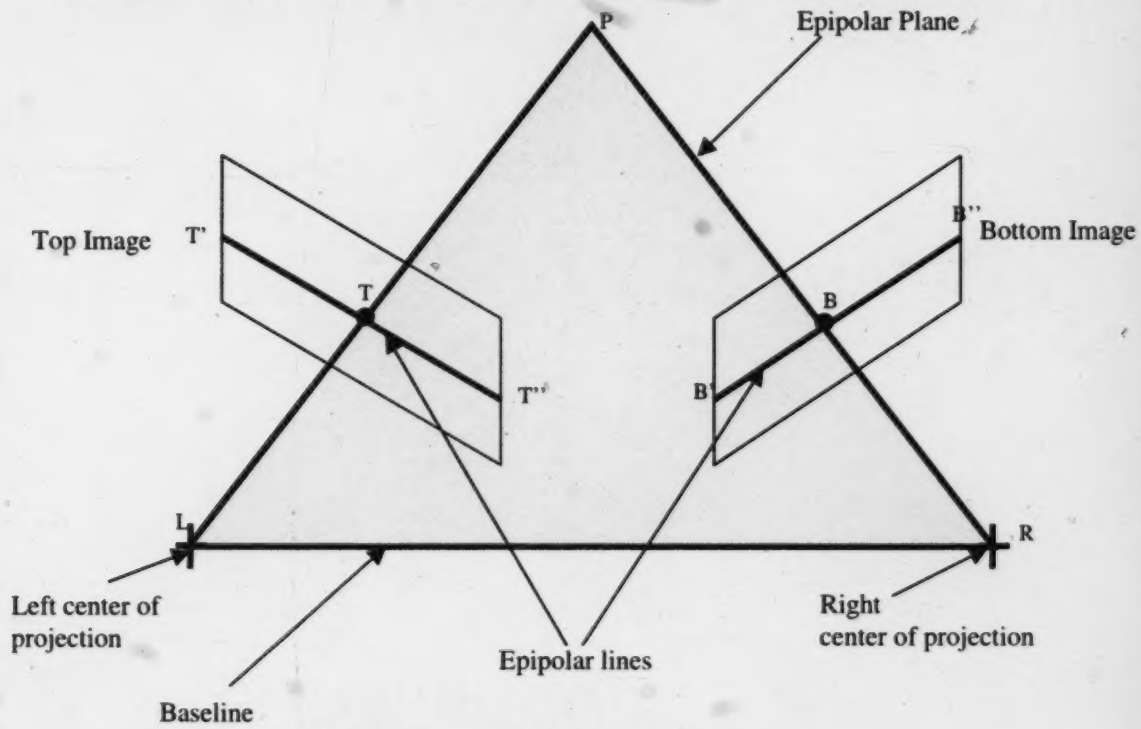
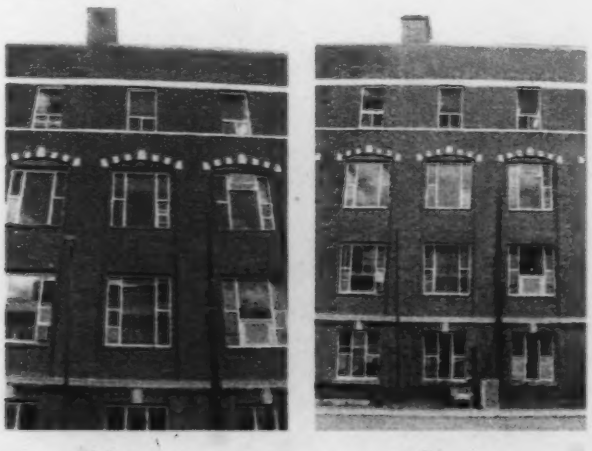


Figure 2.4: General epipolar geometry

That is, when the pair of stereo images is rectified the epipolar lines are horizontal scanlines. A pair of corresponding edges should be searched only within the same horizontal scanline. As illustrated in figure 2.4 given an image point T, its corresponding point (B) in the other image is constrained to lie on the straight line that is the projection

of the line through the given image point and its center of projection. It is constrained to lie on the projection of only that portion of the line that extends outward from the given image point, rather than on the projection of the whole line. The line (LR) connecting the two centers of projection is called the baseline. A plane through the baseline and point P is termed an epipolar plane.

Any such plane will, in general intersect the two image planes along straight lines; these straight lines are called epipolar lines. Any point on an epipolar line (T'T'') has its corresponding image, if any on the corresponding epipolar line (B'B''). This restriction is called the epipolar constraint[29], [45]. Restricting the correspondence search to conjugate epipolar lines results in a great computational saving, over and above the increase in the likelihood of obtaining correct matches. In general, the epipolar lines in each image converge toward the intersection of the image plane with the baseline. For computational convenience, the two image planes are often chosen to be coplanar and parallel to their baseline. Epipolar lines in rectified stereo images are parallel, and conjugate epipolar lines in rectified stereo images are collinear. The images are said to be rectified when the stereo images are coplanar and parallel to their baseline. Figure 2.5 illustrates the concept of image rectification. Figure 2.5(a) shows an image of a building taken with the camera tilted. Figure 2.5(b) shows the image taken with the camera more or less parallel to the building. Rectification is the mathematical procedure of transforming image 2.5(a) to image 2.5(b).



(a) (b)

Figure 2.5: Rectification of an image (*Courtesy: Prof. A.E.Peterson, Univ. of Alberta*).

2.2.4 Correspondence Matching

The difficulty with geometric stereo photogrammetry is the establishment of correspondence, that is, the pairing of points in the two images such that each point in a pair of points is the image of the same point in space. Ambiguous correspondence between points in the two images will lead to incorrect 3D information. The problem of correspondence establishment is further confounded by the fact that, in general, some points in each image will have no corresponding points in the other image. There are two reasons for this absence of corresponding points. First, clearly the two cameras will have somewhat different fields of view, secondly, objects in the scene may cause occlusion of different regions in the two images. There exist two general types of stereo matching, namely, intensity based (area-based) matching and feature based matching. The intensity-based approach is based on point to point matching of image brightness or matching a surface area in one image with patches in the other image. Feature based algorithms compare specific characteristics extracted from a pair of images, such as edges, lines, vertices and other regular shapes. Through this process a best match between two images can be obtained. In general, the correlation algorithms provide good matches, by taking advantage of their inherent noise suppression effects if distortion is not excessive [4].

Many other techniques for stereo matching have been proposed [29] [40] [42] [29] [26]. Marr and Poggio [28] and Grimson [18] proposed an edge based stereo matching algorithm, in which edge pixels that have the same edge polarity, approximately the same orientation and lie on the same epipolar line in the left image are matched with the edge pixels in the right image. Chou and Tsai [9] presented a new iteration scheme for solving the line segment-matching problem, in which a match function directly reflects the requirements of the epipolar and disparity constraints. In Ohta and Kanade[32], stereo matching is treated as a search problem, including the intra-scanline search and the inter-scanline search. The intra-scan line search is to find a matching path on a 2-D search plane whose axes are the right and left scanlines. Vertically connected edges in the images provide consistency constraints across the 2-D search planes, which are used in the inter-scanline search in a 3-D search space. Their algorithm uses edge-delimited intervals as elements to be matched, and employs the

above two searches using dynamic programming. In this research, the stereo matching algorithm will be implemented based on the feature extraction technique using image correlation and the epipolar constraint will be used to match the extracted features. An introduction to feature extraction by image correlation is presented in the following section.

2.2.5 Feature Extraction by Template Matching

Template matching is a simple filtering method of detecting a particular feature in an image. Provided that the appearance of this feature in the image is known accurately, one can try to detect it with a template operator. This template is a subimage that looks just like the image of the object. A similarity measure is computed which reflects how well the image data match the template for each possible template location, in terms of size, shape and relative intensity. The point of maximal match can be selected as the location of the feature. The basic idea of template matching is to find an area of high correlation between a template (a fixed 2-D pixel pattern) and a subsystem of the same scale in the image. The correlation is done on a pixel by pixel level and is measured by the number of matched pixels.

Image Correlation

Consider a subimage w(x,y) of size J x K within an image f(x,y) of size M x N, where we assume that J < M and K < N (figure2.6). In its simplest form, the correlation c(s,t) between f(x,y) and w(x,y) is

$$c(s,t) = \sum_x \sum_y f(x,y)w(x-s,y-t) \quad (2.1)$$

Where s = 0, 1,2.....M-1, t= 0,1,2,.....N-1, and the summation is taken over the image region where w and f overlap. The maximum value of c(s,t) indicates the position where w(x,y) best matches f(x,y).

The correlation function in equation 2.1 has the disadvantage of being sensitive to changes in the amplitude of f(x,y) and w(x,y).

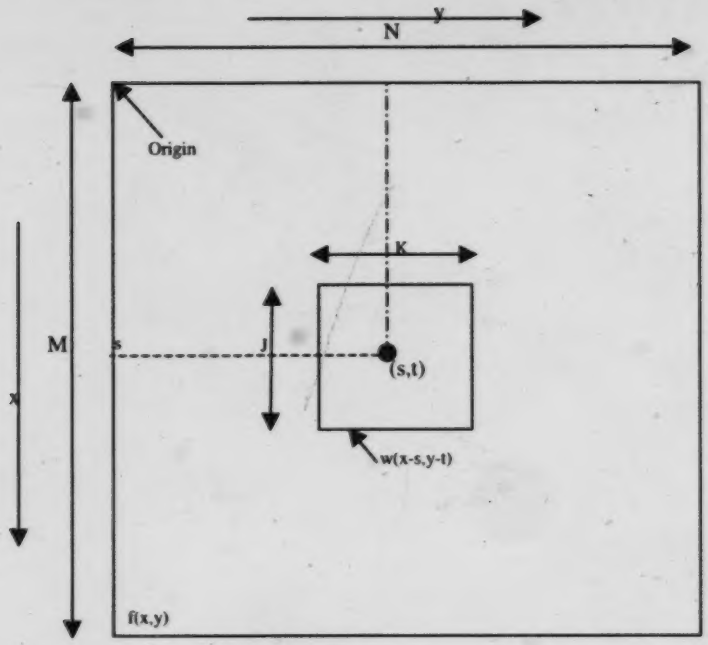


Figure 2.6: Arrangement for obtaining the correlation of f(x,y) and w(x,y) at point (s,t)

An approach frequently used to overcome this difficulty is to perform matching via the correlation coefficient, which is defined as

$$\gamma(s,t) = \frac{\sum_x \sum_y [f(x,y) - \bar{f}(x,y)][w(x-s,y-t) - \bar{w}]}{\left\{\sum_x \sum_y [f(x,y) - \bar{f}(x,y)^2 \sum_x \sum_y [w(x-s,y-t) - \bar{w}]^2\right\}^{1/2}} \qquad (2.2)$$

Where s= 0,1,2,.....M-1,t=0,1,2.....N-1, $\bar{w}$ is the average value of the pixels in w(x,y) (computed only once), $\bar{f}(x,y)$ is the average value of f(x,y) in the region coincident with the current location of w, and the correlation coefficient γ(s,t) is scaled in the range –1 to 1, independent of the scale changes in the amplitude of f(x,y) and w(x,y).

Correlation can also be carried out in the frequency domain using the Fast Fourier Transform(FFT). If f and w were the same size, this approach can be more efficient than

direct implementation of correlation in the spatial domain. If equation 2.1 is used, w is usually much smaller than f, a trade-off estimate proposed by Cambell [16] indicates that, if the number of nonzero terms in w is less than 132 (a subimage of approximately 13 x13 pixels), direct implementation of equation 2.1 is much more efficient than the FFT approach. This number, also depends on the machine and algorithms involved in the implementation, however it does indicate the approximate subimage size at which the frequency domain should be considered as an alternative. The correlation coefficient is considerably more difficult to implement in the frequency domain and is usually computed directly from equation 2.2. The feature extraction process using the template matching technique is discussed in detail in Chapter 4.

2.3 Literature Survey of Coding Schemes

This section surveys the structured light techniques for 3-D surface reconstruction. An attempt has been made to classify and evaluate the available measurement methods and thereby identify the techniques that can be used for recovering the three-dimensional coordinate information of a human trunk. These techniques are investigated further by conducting tests to determine if they can be applied within the constraint of the present stereo vision system. The techniques have been classified as binary, three level and colour coding schemes. The following subsections discuss variations in each scheme proposed by researchers.

2.3.1 Binary Coding

This scheme of coding uses only two levels of illumination dark (black) and bright (white) levels. A number of projection patterns have been proposed using binary coding schemes [10], [14], [22], [23], [31], [43].

The binary projection pattern can be points, line segments, or more complicated patterns such as parallel sets of stripes and orthogonal grids. The idea of using spatially modulated light patterns projected onto a scene for analyzing was first proposed by Will and Pennington [44]. An orthogonal grid pattern or a set of parallel stripes was projected to encode planar surfaces. In general, distortion of the projected patterns on a planar surface is a function of the surface orientation and the position in space, with respect to

the projection plane and the image plane. The distortion of the projected patterns manifests itself in an angular and spatial frequency carrier shift in the frequency domain. Therefore, the extraction of planar regions from the image becomes a matter of linear frequency-domain filtering. Thus grid coding provides an alternative to other heuristic approaches for image segmentation and makes use of the extra degree of freedom (the nature of scene illumination).

Will and Pennington have also proposed a stereo setup using grid coding for constructing three-dimensional information automatically

In general, techniques developed for inferring three-dimensional structures from the observed spatial encoding proceed in three steps:

1) The relative position of the controlled light source and the image plane is determined through a registration procedure.
2) The correspondence relations of features (for example, the grid junctions) in the grid plane and the image plane are then established, and the spatial positions of features are recovered using triangulation and back projection.
3) The surface structure in three-dimensional space is determined by interpolating the sparse depth map of features.

However, using grid coding for depth reconstruction using the stereopsis principle poses some difficulties. It is difficult if not impossible to find correspondence between the grid intersections in a stereo pair of grid coded images. Matching features extracted from different image frames for the recovery of depth is still a nontrivial undertaking, because the evaluated features are numerous and indistinguishable. Heuristic search [36], time modulation [35], spatial labelling [30], relaxation [21] and colour encoding [2] have been proposed to alleviate this problem of correspondence.

Heuristic Search

Potmesil and Freeman have adopted a heuristic search algorithm for determining the correspondence between grid junctions in different images [36]. They presented a method for constructing surface models of 3-D objects by spatially matching 3- surface segments describing the objects. The object was placed on top of a rotating turntable to reveal different portions of the object surface to the cameras. The surface of the object to

be represented is imaged from multiple stable positions. Furthermore, multiple cameras may be used for each turn table position, which gives much freedom in image capturing. A calibration backdrop is carefully assembled for registering different camera positions automatically. In this method projected grid patterns are used to recover the surface structure from multiple viewpoints. Calibration marks are placed in known positions on the backdrop. The positions of the cameras with respect to the object are calculated automatically from the perceived configuration of the calibration marks. Once the relative position of the cameras is obtained through the automatic calibration process, a heuristic search algorithm is applied to establish the correspondence relationship of the grid junctions between image frames. The three-dimensional positions of the grid junctions are then recovered through triangulation, and the descriptions of the visible surfaces are obtained through bi-cubic spline fitting. A heuristic search algorithm is used to find the spatial transformation between any two overlapped surface elements by maximizing the similarities in their position, orientation, and curvature measurements. The authors also report that a complete description of the surface can be generated by the use of relatively large number of grid coded images. However, the heuristic search is usually a time consuming process.

Spatial labelling

Another possibility in assigning a unique space coding to the grid junctions is through the use of a spatial labelling technique. Moigne and Waxman [30] proposed a grid of horizontal and vertical lines with several dots. The black dots are detected in the image, and a counting procedure that measures the offset of a junction to the black dots is used to determine the space code of the junction. The labelling process then proceeds to locate the registration dots and measures the displacement of a junction to the dots. Note that discontinuities of the grid lines and missing intersections will mislead the address labelling process. Hence, a global relaxation process is incorporated to resolve the inconsistency in labelling. However this spatial labelling fails when abrupt changes in depth cause a break of the observed stripe patterns because the grid intersections are used directly in stereo matching.

Lavoie and Petriu [25] proposed a method for determining a set of reference pixels in two simultaneous views of the same object, using two cameras, by projecting a

pseudo-random encoded grid on the object. The grid nodes and their encoding values are extracted from 2-D images by applying first a smoothing and then a watershed algorithm. This method offers three distinctive advantages over a conventional stereo system:

- easily generates a list of matching points.
- adds structure to an otherwise texture-less object.
- is less computational intensive.

When the grid is encoded with a pseudo-random binary array (PRBA) (as shown in figure 2.7) a correspondence can be made, since each small array of size k1 X k2 is unique in the PRBA; by simply looking at the pattern present in the small array the exact position from the PRBA can be deduced.

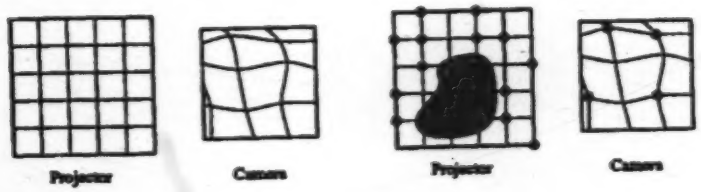


Figure 2.7: When grid is not encoded When grid encoded with dots to form PRBA

A pseudo-random binary array (PRBA) is defined by a n1 by n2 binary array encoded using a pseudo-random sequence such that a k1 by k2 window sliding over the array is unique and fully identifies the window's absolute coordinate (i,j) within the array. The following relations hold for the PRBA,

$$2^n - 1 = 2^{k1k2} - 1 \qquad (1)$$

$$n1 = 2^{k1} - 1 \qquad (2)$$

$$n2 = (2^n - 1)/n1 \qquad (3)$$

where n1 and n2 must be relatively prime. (Numbers are relatively prime if their only common divisor is 1. For example 6 and 35 are relatively prime.)

The process by which the "primitive polynomials modulo 2n" method generates random bits from a primitive polynomial is described by Press et al [37]

Square Checker Board Grid

Vuylsteke and Oosterlinck [41] developed a system which projects a square checkerboard grid (figure 2.8) of black and white points onto the scene which is a binary-encoded light pattern (figure2.9).

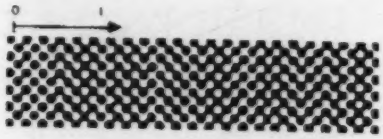


Figure 2.8: Projected binary pattern

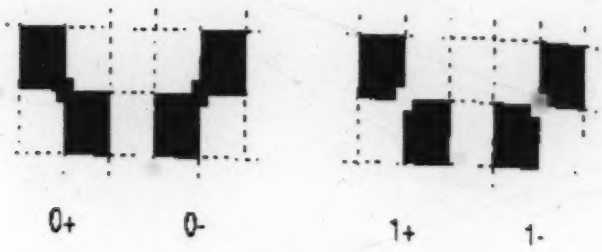


Figure 2.9: The Illumination pattern is composed of four primitives

A binary signature is defined for each point by looking at the values of the neighbors in a window about that point. The encoded light pattern is generated from pseudonoise sequences (which generate the 0 and 1 values for the pattern recursively to define which points are black and which points are white). The authors optimize the pattern in the sense that the pattern is fault tolerant to the largest extent. That is, a minimal code distance, d, may be specified between non-unique signatures. Since a binary signature is defined in a small window, this method is not susceptible to occlusion as in the method proposed by Boyer and Kak [47]. However for a general m x n pattern it is still possible for multiple correspondences to exist since the pseudonoise sequences do not guarantee uniqueness, but only that non-unique signatures be at least a distance d apart. Further, complete information of the coding primitives is not used. Finally for a large binary pattern, large windows must be searched for the binary signatures.

The authors have stated that one grid point corresponds to an average 8 x 8 pixels, the input picture being 512 x 512 pixels. The ratio of the one grid point to the image size

is 8:512. The lateral resolution of the range image can be improved by reducing this ratio (8:512) to 4:512.

2.3.2 Three-level Coding

The projection of lines can be regarded as light sectioning planes, labelling the measuring space. As the projected information is to be extracted from the image, every section needs a different label. By increasing the number of labels the measuring volume can be extended and the resolution can be improved.

The labelling of the sections occurs primarily by the intensity of the illumination. If the projecting unit is capable of producing 3 different intensities of light in a single image, 9 different sections can clearly be labeled (figure 2.10).

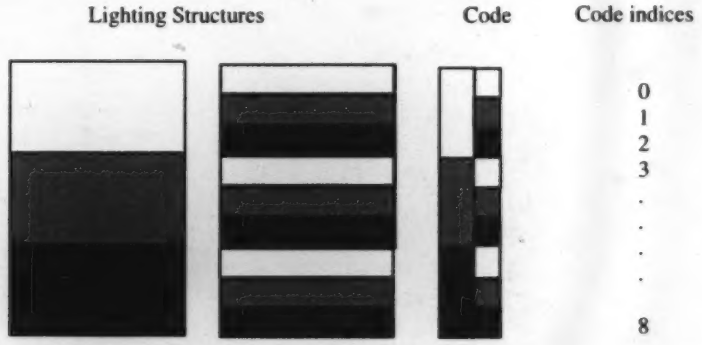


Figure 2.10: Coding the lighting sections

Another means of labelling the direction of projection relies upon the sequential projection of several different light patterns on the object (figure 2.11). The camera records each of the patterns projected onto the object.

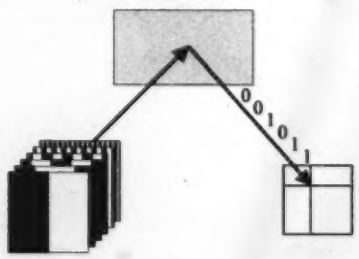


Figure 2.11: Binary line coding using multiple patterns

The projection unit produces a sequence of intensities for each lighting section. The extraction of the gray value of each pixel position in the different recordings yields a sequence of gray values for each pixel [20]. The sequence of light intensity values can be regarded as a code, corresponding to a specific lighting section.

The phase shift method uses another kind of gray-coded illumination. Three sinusoidal fringe patterns, shown in figure 2.12 are projected onto the object sequentially the second sinusoid 90^{o} with respect to the first shift the third sinusoid 90^{o} with respect to the second.

Figure 2.12: Three sinusoidal fringe patterns for phase shift illumination

From the detected phase shifts in the viewed gray scale image the three dimensional object range could be calculated. However, the periodicity of the coded illumination causes ambiguities in the calculated range map.

All these represented methods require more than one snapshot for the complete object acquisition. Patient movement from one snapshot to another can affect the measurements significantly. To avoid motion artifacts the camera should sample the scene at about 30 frames per second.

The colour-coded phase shift method has been proposed by Schubert et al [39]. It is based on the gray-scale phase shift method. The three required sinusoidal illuminations of the gray coded phase shift method are coded in red, green and blue and they are combined to one colour-coded illumination. Using this illumination, the acquisition time compared to the gray coded phase-shift method is reduced to one third. Compared to the colour triangulation, with the colour-coded phase shift method a higher spatial resolution could be realized, but still the ambiguity is present.

2.3.3 Colour Coding

In the scheme proposed by Boyer and Kak [2] a single coloured stripe pattern is projected on the scene. Only a few different colours are used, typically three or four. The illumination pattern can be thought of as a concatenation of identifiable subpatterns (typically six stripes wide). The recognition of any valid subpattern codeword in the observed stripe sequence leads to the identification of the stripes in that subpattern, and possibly also of some adjacent stripes (by some kind of crystal growing process).

A colour video camera or a black and white camera with a set of suitable colour filters is used to generate images in different visual spectrum bands. Preprocessing modules are incorporated to remove noise in the high and low frequency bands and to equalize the signal strength in different colour images.

The stripe indexing technique assigns a unique address to each observed colour stripe. Corresponding scan lines in different colour images are analyzed. A peak detection module detects the existence and location of stripes in different colour images. A white stripe is identified by the simultaneous presence at the same scan position of stripes in all colour images. The received pattern is then assembled in the line pattern format module. The received pattern is made of a list of positions of the detected stripes in the current scan together with the colours of stripes.

The transmitted pattern is composed of subpatterns of distinct arrangements of parallel colour bars. The received pattern is matched against the transmitted subpatterns to find perfectly matched positions in the received pattern. These perfect-match positions serve as "seeds" to generate larger matched segments, or the matches are expected over the neighborhoods. Possible ambiguities are in the assignments of correspondences are then resolved by a 'survival of fittest' operation .

Notice that each transmitted stripe can be imaged only at one position, also any received stripe results from the projection of one and only one, transmitted stripe. However, since this encoding scheme is based on adjacent subpatterns, small amounts of occlusion can greatly affect the performance.

Griffin and Narasimhan proposed a method for generating an encoded pattern which may be used for a special structured lighting system [17]. This pattern consists of a

matrix of coloured circles which is projected onto a scene by backlighting, as shown in figure 2.13. A single camera is used to image the projected colour coloured light beams, and the range data for the object is obtained. Since the light pattern is encoded, the correspondence between the projected beams from the pattern and imaged beams is easily determined.

Further the authors have also mentioned that this method is not susceptible to problems with occlusion and it is not restricted to binary patterns, any size pattern can be generated and thirdly it is guaranteed that each position defined in the pattern is unique.

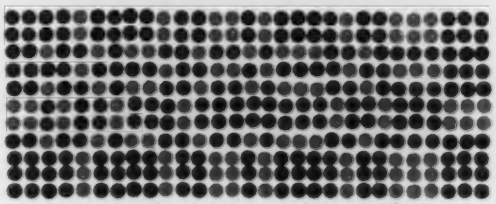


Figure 2.13: The pattern used for the coloured lighting system with the colours red, green and blue.

However it is possible to make false colour readings based on the colour content in the scene. This approach works best for environments that are colour neutral. In many applications, the scene may not be colour neutral or the vision system used may only have gray-scale capability. In this case, the structured light must be monochromatic. The same authors have also proposed a possible encoding scheme using a modified grid pattern where a primitive is assigned to each grid intersection point as shown in figure 2.14

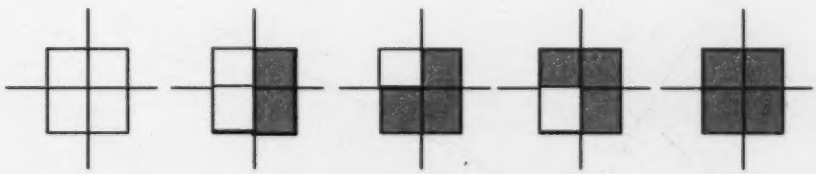


Figure 2.14 The five modified grid primitives

Patterns are generated by assigning colour primitives to elements of a matrix M. A word of M is defined by the colour at location x_{ij} and its adjacent neighbors. The primitives are assigned in such a way that no two words in M are the same.

2.3.4 Summary of Literature Review

A number of coding techniques were reviewed. Each approach has its strengths and drawbacks, and is suitable for certain classes of objects, or certain environments. Some of the complex patterns described above cannot be used in the current system due to the resolving capabilities of the camera-frame grabber system, and such coding schemes also require high precision projecting devices. The technique of projecting multiple patterns and analyzing each frame cannot be applied to human subjects because the effect of breathing artifacts and other patient movements can affect the performance of the system.

Horizontal black and white lines have been used previously for the reconstruction of the trunk surface [27]. The vertical resolution achieved was 13mm. Higher resolution can be achieved by using densely encoded patterns and extracting the features efficiently. However this requires a coding scheme which has distinct features that can be easily identified and matched with certainty in the other image. Multiple levels of illumination is therefore one of the promising coding schemes that can be tested for this system. Colour coding scheme is also another technique that can be useful in this application, however this requires colour cameras.

From the various schemes that researchers have used in the past, three classes of spatial encoding schemes show promise. They are the binary, three-level, and multilevel or colour coding schemes. These techniques deserve further investigation to determine which is the best-suited coding scheme for the system used in this research. To be eligible as an efficient coding scheme, the pattern must fulfil the following requirements: the code has to be uniquely defined and the code should have low spatial frequency. High spatial frequencies make a measurement sensitive to defocusing and to steep rises of the surface and should therefore be avoided.

The resolution of the cameras has to be studied to establish the feasibility of using complex coding schemes that have high spatial frequency.

To investigate these techniques experiments have to be carried out using slides with each of the coding schemes to be tested. The tests and results obtained are discussed in detail in the following chapter.

3 Experiments to identify an appropriate coding scheme

The association of sequences of intensities to light sections, called coding, plays a major role in range data acquisition by structured light methods. The coding scheme has an influence on the resolution, the speed and the robustness of the measurement. The objective of this research is to identify a coding scheme that aids in solving the correspondence problem and combines high resolution with short measuring time.

An efficient coding scheme is uniquely defined. This means that each lighting section obtains a unique label for identification. A code with low spatial frequencies is more robust because there are no decoding errors due to the integration of a camera pixel. High spatial frequencies make the measurement sensitive to defocusing and to sudden depth changes on the surface and therefore should be avoided. The code should use a maximum of the available codewords (primitives). The computational time involved in extracting the coded features from the image should be clinically acceptable. A clinically relevant time is about 5-10 minutes. A high number of codewords is required because it increases the resolution of the measurement. With lower number of codewords decoding error occurs and also the resolution decreases.

Considering the above requirements, binary, three level and colour coding schemes were tested and the effectiveness of each scheme was evaluated. This chapter reviews the experiments conducted, to identify an appropriate coding scheme that can be used to perform 3-D surface measurements of the human trunk. The ability of each scheme in solving the correspondence problem is discussed in separate sections. From the results obtained the appropriate coding scheme will be identified. This pattern will be projected on the surface to be measured and image processing techniques will be used to extract the features from the image.

3.1 Experimental Setup

The system (figure 2.3) consisted of a pair of Sanyo VDC-2524 cameras, Tamron variable focal length (6-16mm) lenses, a Scion II image capture Card used with a Macintosh computer to capture images. A Kodak projector was used for displaying the 640 x 480 pixels pattern onto the subject, such that it covers a region of 640mm by

480mm. As shown in figure 2.3 the cameras were arranged with a vertical baseline of about 1000mm. The cameras were aligned vertically so that the baseline between the two cameras is parallel to an axis of the cameras, then the correspondence needs to be searched only along lines parallel to that axis in the image. By constraining the geometry of the cameras along one axis, it is sufficient to restrict the coding scheme to one direction (vertical direction in this case). Applying the epipolar constraint after rectifying the stereo images solves the correspondence problem along the other direction. To carry out the experiments the cameras were calibrated and slides created with different coding schemes.

3.1.1 Calibration Procedure

To calibrate both cameras in the system, a control frame was positioned vertically on the object stand. The control frame consists of thirty control points. Their world X, Y,Z coordinates with respect to the origin are known [27]. The control frame images from both cameras are recorded. Then a one to one manual correspondence matching is performed on the x-y image coordinates of the 30 control points. Since the world three-dimensional coordinates of each control point is known, and the correspondence matching is done manually, the exterior and interior parameters of the cameras were computed using the calibration program [34]. The collinearity equations and the least squares solution method are used in this program. By fixing the second radial distortion parameter, K2 and the decentering distortion parameters, P1 and P2 to zero, the solutions of the system of equations converged within 5 passes. Nominal values were assumed for principal point co-ordinates xp and yp. The following data (table 3.1) was obtained and these parameters are used in testing the system.

Cam	X (mm)	Y (mm)	Z (mm)	ω (deg)	ϕ (deg)	κ (deg)	K1	Focal length(mm)
1	-254.3	317.2	1477.5	-0.37	-18.03	179.2	-0.18E-3	15.211
2	794.1	315.1	1463.6	1.40	16.48	179.4	-0.18E-3	14.919

Table 3.1: Results of camera calibration

3.1.2 Projector Slide Design

To create a slide with the desired pattern, an image has to be created which can be printed onto a slide. To create an image a C program was written which would convert a data file into a tiff image with the desired pattern. This program consists of two parts. The first part creates a data file, which would contain the coded pattern in the form of an array of gray level values. The second part converts the image array to tiff image format. This tiff image is then printed onto the slide and the projector is arranged in such a way that 1 pixel on the slide covers approximately 1mm on the surface of the subject. This set-up is used to create the codes that were selected for further investigation.

3.2 Binary (Two level) Coding Scheme

In this section, the properties of the binary coding scheme and its performance are evaluated. This method of coding requires only two different intensities. A projected line is either 'on' or 'off'. Only maximum and minimum illuminations are used for labelling the different illumination sections. This is represented in the figures by black and white lines.

The advantage of using black and white line patterns as projected light is that the image processing steps in thresholding are simpler. Since there are no different levels of gray in the image, a certain pixel can be forced to white or black depending on a threshold value.

Previous experiments indicated a few problems in using just black and white lines [27]. One of the drawbacks of using horizontal parallel black lines is that there is no definite way of tracing (identifying) these lines when a sudden depth change occurs. Figure 3.1 shows an object that has a depth change of 5cm, 3 cm and 1 cm. When a set of parallel lines are projected onto it, the change in depth causes the lines to shift. This shifting of lines makes it difficult to trace the lines.

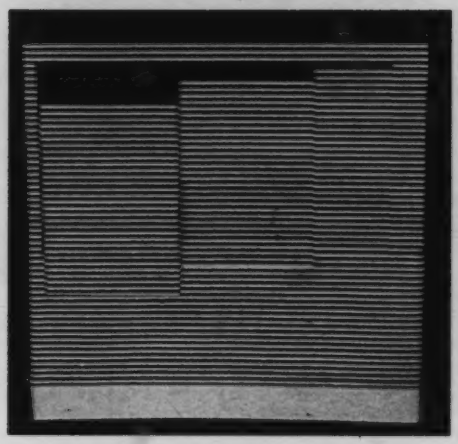


Figure 3.1: Step object with horizontal lines

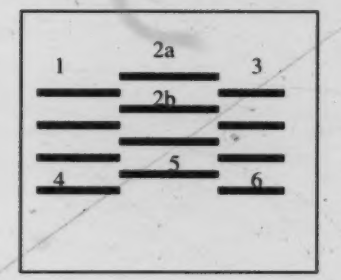


Figure 3.2: Shift of lines due to steep changes in depth

When the subject has a depth change in the direction of the projected light the lines appear as illustrated in figure 3.2. In this situation it becomes difficult to trace the lines. If the lines are followed from left to right, starting with the first line, it is not possible decide whether line 1 is continued by line segment 2a or 2b. However if we started tracing the lines from the last line in the image it is evident that line 4 can be traced along line segment 5 followed by 6. This imposes the constraint that the number of lines in both images (stereo) is the same. However in the present camera set up, the number of lines in both images may not be the same due to occlusion. Occlusion is a phenomenon where a point in one scene is hidden in the other. This is illustrated in figure 3.3, where the stereo images from the top and bottom camera have certain points which are not in the view of the other camera due to the geometric arrangement of the cameras and the shape of the object itself. This results in a number of lines that cannot be matched since they appear in one image and are hidden in the other.

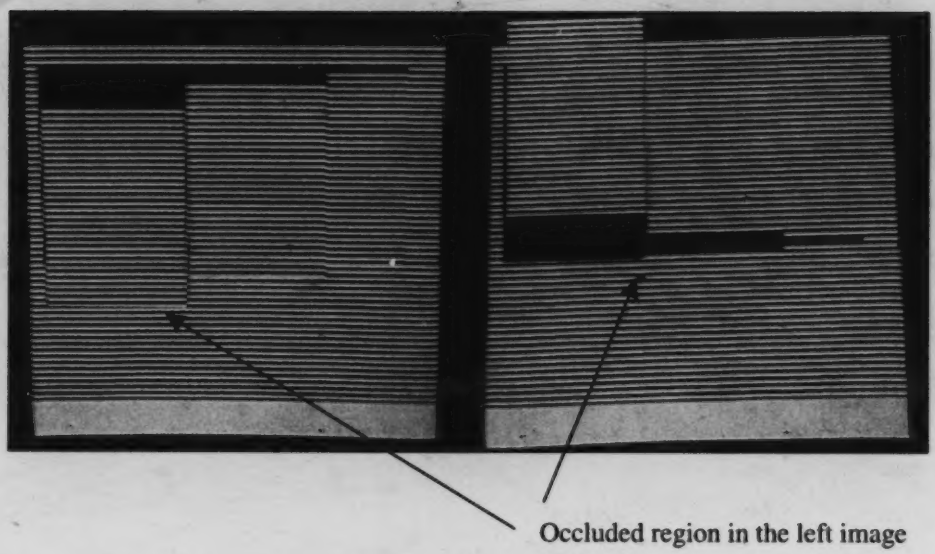


Figure 3.3: Shows Occlusion

This problem can be approached using lines of different thickness. However, this greatly reduces the number of lines that can be projected onto the object. Lowering the density of the lines lowers the resolution of the measurement. Another approach to this problem is to add vertical lines to identify the intersections (figure 3.4). Grids of lines carry more coding information but do not necessarily give more depth information. If a pattern of light stripes can be projected so that they are parallel to the horizontal axis in the observed image, then there is no point in projecting another set of lines parallel to the vertical axis (since these merely replicate the information that is already available from the columns of pixels in the image). The horizontal lines and their vertical displacements in the image carry all the depth information.

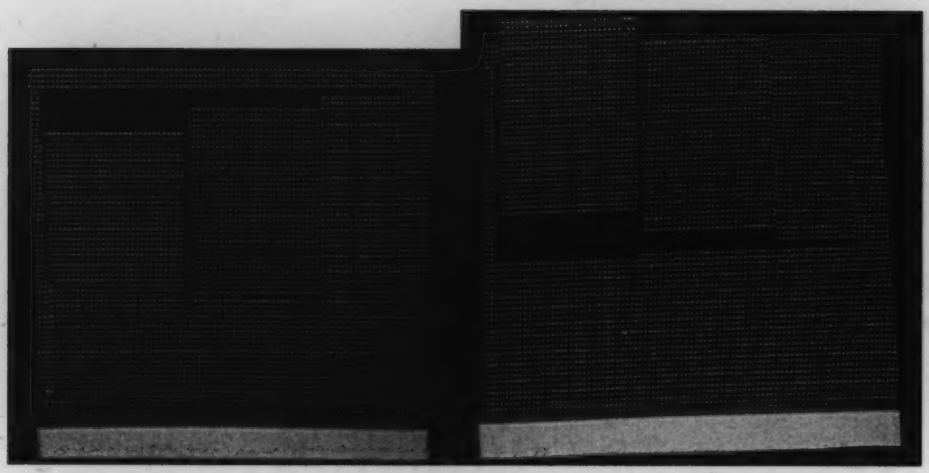


Figure 3.4: Vertical and horizontal lines projected on an object

Another drawback of the present camera-projector system is the distorting effect of the projector and the frame grabber on the captured images. To study this effect slides with black horizontal lines were created. Each slide was created with lines varying from 2 to 6 pixels wide. These were projected on a plane surface. Figure 3.5 shows the slides with the projected 2 pixel wide lines. The present camera system was not able to identify lines, which were 2 pixel thick. This is because the projector brightens the image to such an extent that the image appears to be predominantly white to the cameras with traces of dark lines as shown in figure 3.6.

A column of the projected image was sectioned out and the intensity level is plotted against the number of pixels along a column. Figure 3.7 shows the plot for the projected image. Figure 3.8 shows the plot for the captured image. The intensity levels for the captured image (Figure 3.8) ranges from 140-255, while the plot of the projected pattern has just two levels, 0 and 255. Further the effect of the projector is evident from the fact that level 0 has been shifted to level 140 by the frame-grabber. This is due to the fact that the thickness of the black lines is not enough for the cameras to identify them as black pixels of level 0. From these plots it is evident that dark lines of two-pixel thickness cannot be identified well. A set of slides with lines 3 pixel thick was created.

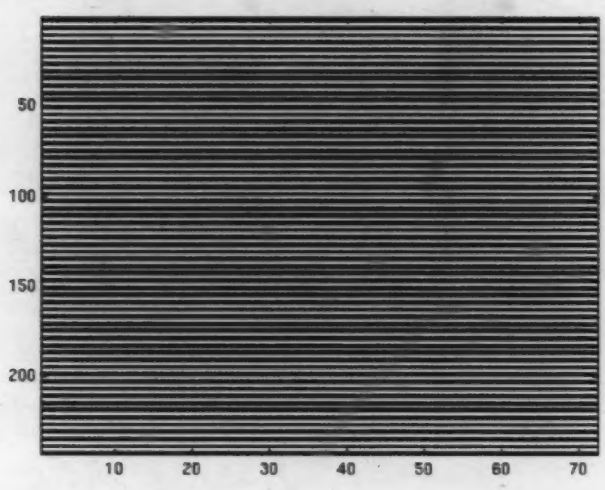


Figure 3.5: Projected slide with two pixel wide lines

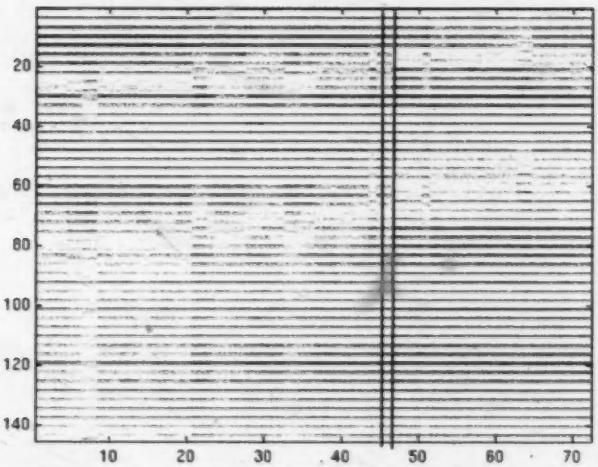


Figure 3.6: Captured image of a 2 pixel wide pattern

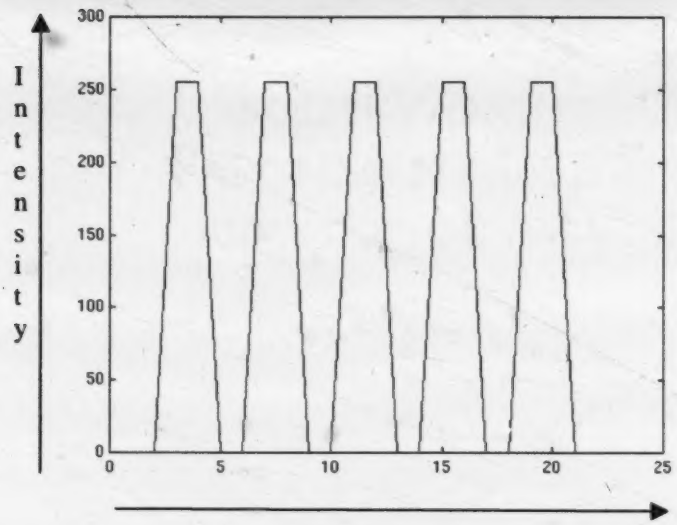


Figure 3.7: Plot of intensity vs number of pixels of projected pattern

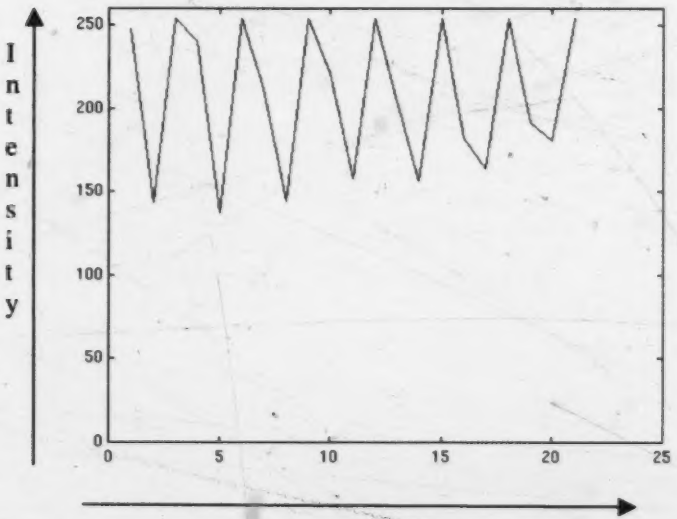


Figure 3.8: Plot of intensity vs number of pixels on the captured image

When these images were captured, it was observed that the line width in the captured image was not consistently 3 pixels because excessive lighting causes the black lines to be thinner than the white lines and the frame grabber integrates the bright areas as white lines. A pattern with 4 pixel thick alternate black and white lines was created (figure 3.9). Figure 3.10 shows the captured pattern and the intensity levels across one column of the projected pattern are shown in figure 3.11. Figure 3.12 shows intensity along a column for the captured image.

From figure 3.12, it can be observed that the thickness of the black lines is less than that of the white lines although the projected pattern has lines of equal thickness. The intensity levels for the captured image (figure 3.10) ranges from 0-255, that conforms with intensity range of the projected pattern. With lines of 4-pixel thickness the intensity levels have a range of 0-255 which is not possible to achieve with lines of two or three pixel thickness. This helps to identify the number of black lines on the image. With the present camera projector system, the minimum thickness of the lines to be projected is 4 pixels.

The CCD cameras are capable of capturing 800 x 500 pixels, however the frame grabber size is 640 by 480. A jitter of ± 1-pixel results if there is no exact synchronization of the camera and frame grabber resolution, and this reduces the resolution of the captured image. This restricts the use of black and white lines of different thickness, since there is no direct relationship between the projected pattern and the captured image. The two level coding scheme can also lead to false line interconnection when projected onto the human trunk, if the densely spaced lines are shifted due to the trunk shape. Hence with the present camera system, two level coding schemes will not increase the previously achieved line density of 13 mm.

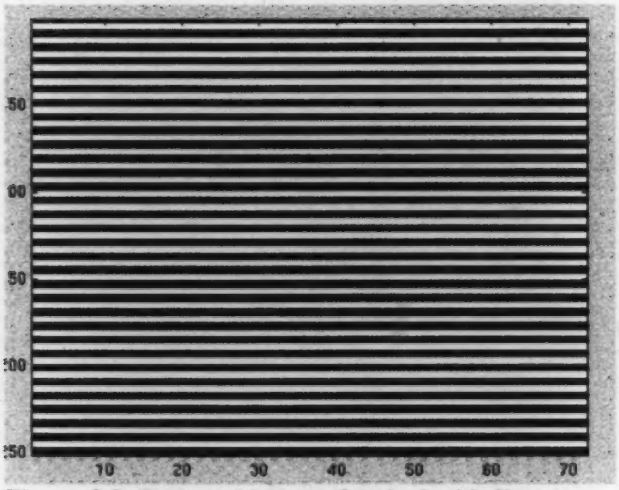


Figure 3.9: Projected image of 4 pixel wide lines

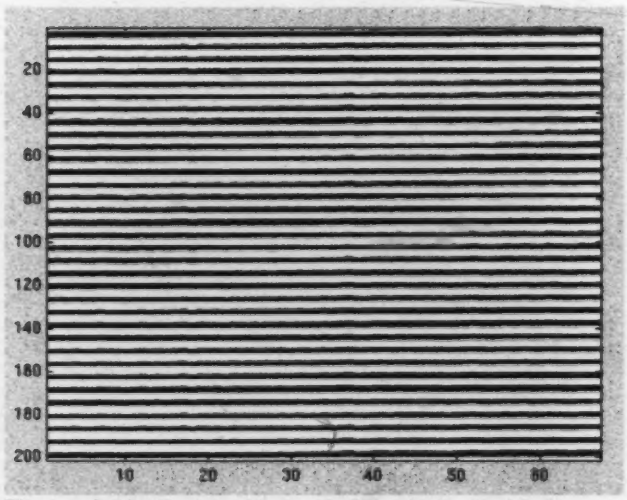


Figure 3.10: Captured image of the 4 pixel wide lines

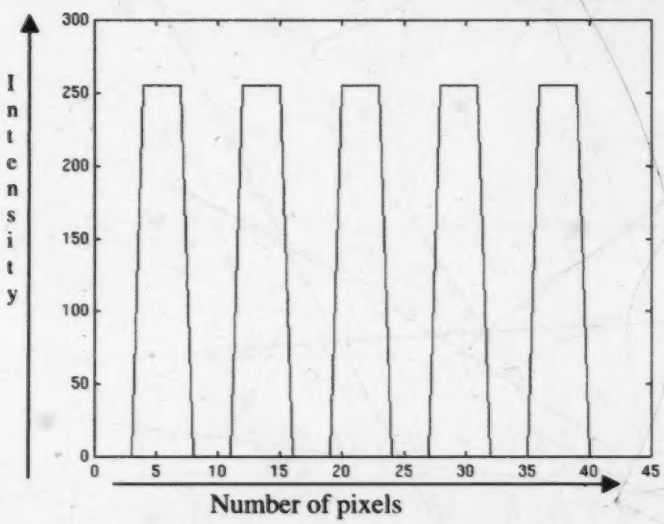


Figure 3.11: Plot of intensity vs number of pixels of the 4 pixel wide projected lines

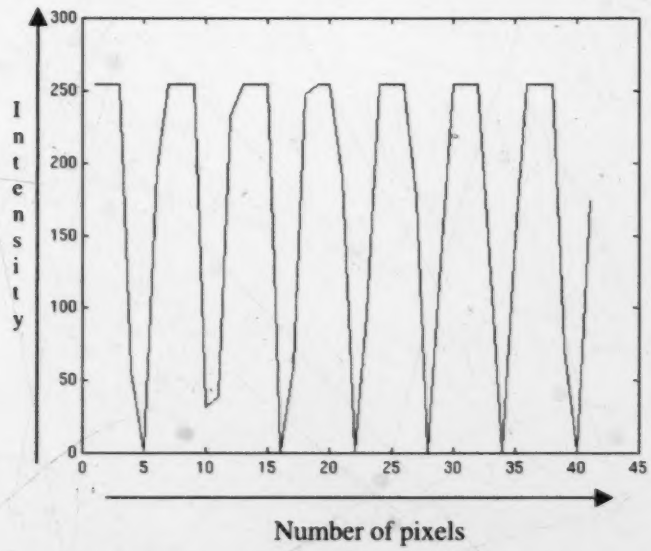


Figure 3.12: Plot of intensity vs number of pixels of the captured image

3.3 Three Level Coding

In this method of coding, three levels of intensity are used for labelling the subject. The three levels are black, white and an intermediate level, gray. This method of coding is advantageous over the binary-coding scheme because the third level, namely gray can be used as a separator between adjacent codes, especially when sudden depth change occurs, tracing the continuation of a broken line becomes easier. Also with three levels, 6 combinations (non-repetitive) can be produced (figure 3.13).

Figure 3.13: Three level coding schemes

The six combinations are BWG, BGW, WBG, WGB, GWB, GBW. These six combinations can be arranged in a number of ways to generate a finite number of sequences. Twenty-three such sequences were chosen and slides with three levels of intensity were created. Previous experiments indicated that the minimum thickness of the lines for the present camera/grabber system should be 3 pixels. Hence these slides were created with gray, black and white lines with a thickness of 4 pixels. Each combination has three lines of 4 pixel thickness. Since there are 6 combinations, the sequence repeats every 72 pixels(6x3x4). Each slide was then projected onto to a plane surface to study the effect of each these patterns on the camera system. Figure 3.14 shows one such projected sequence. The captured image is shown in figure 3.15. The intensity level plots for the projected and captured images are shown in figures 3.16 and 3.17 respectively.

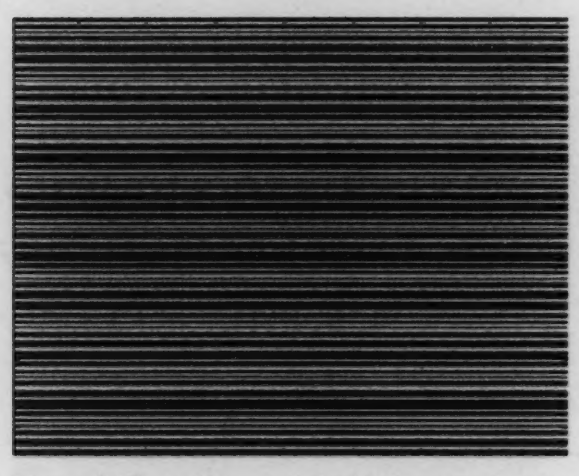


Figure 3.14: A slide created with three level coding

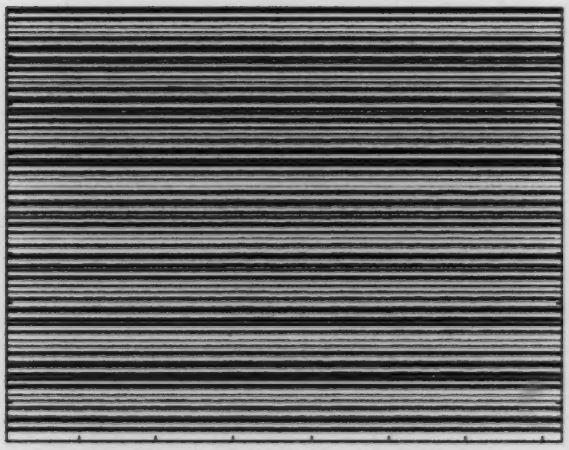


Figure 3.15. Captured image of the three level pattern

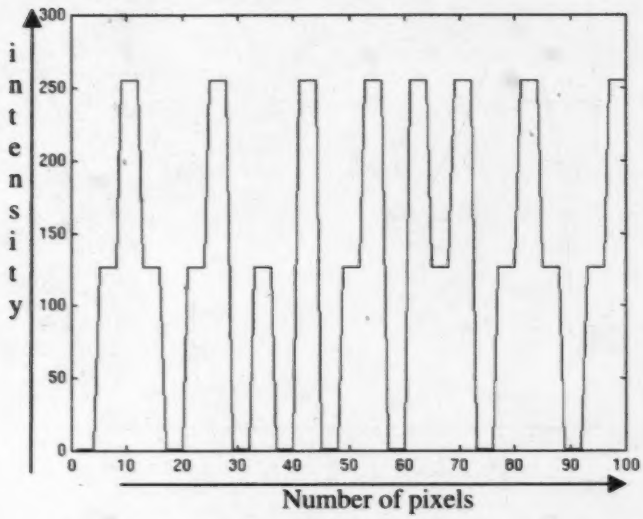


Figure 3.16: Plot of intensity vs number of pixels of the projected pattern

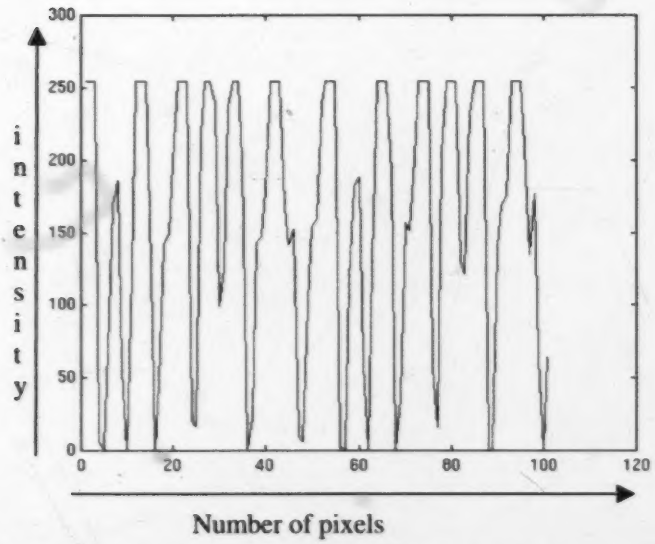


Figure 3.17: Plot of intensity vs number of pixels of the captured image

From the plot (figure 3.17) it was found that the thickness of each line varies with the brightness of the adjacent line. However the three distinct levels can be identified by template matching. Therefore with a line width of 4 pixels, and known sequence of codes the present system can reproduce an image whose features can be identified by an 'intelligent' matching algorithm.

To identify the combination that would enable more features to be extracted accurately, a set of slides with 8 different combinations of gray, white and black lines were created. All these lines were 4 pixels thick. These were tested by projecting onto the plane surface. Two significant factors were observed. One was the effect of the projector and the second effect was that of the frame grabber. The projector's brightness greatly reduced the width of the gray line whenever it was sandwiched between two white lines, more than when it was between two black lines. The plots in Figure 3.18 illustrate this phenomenon. From the plots it can be observed that the number of gray pixels in figure 3.18a is two while the actual number of pixels projected is four (figure 3.18b) .

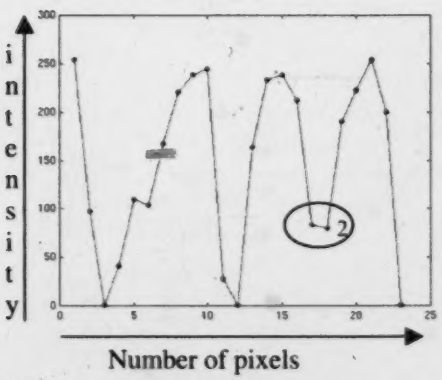


Figure 3.18a: Detected levels

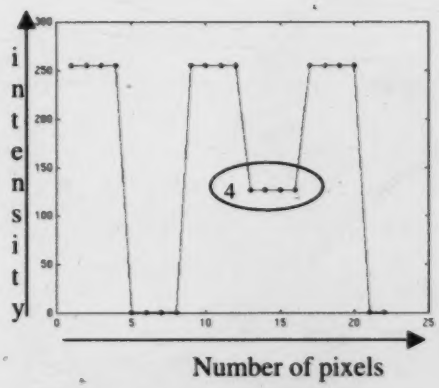


Figure 3.18b: Projected levels

Secondly the framegrabber would distribute the intensities among the 256 levels. The captured image would not have lines of 3 intensities and equal thickness but instead have 256 levels of different thickness. The first problem is significant because the code itself changes prior to image capture. This would defeat the purpose of applying a known

pattern of light. Further tests were conducted in chosing a pattern which would minimise the distorting effect of the projector on the known pattern. One of the schemes is to avoid having a gray line in between two white lines. This scheme restricts the use of some of the sequences that can be created with 6 combinations Restricting the number of combination increases the repetition rate of the sequence. Some of the sequences that cannot be used are:

WBG BGW BWG WGB GBW GWB

WBG BWG BGW GBW GWB WGB

The sequences that can be used are:

1. GWB WBG WGB GBW BWG BGW
2. GWB WGB WBG BWG BGW GBW
3. WBG WGB GWB GBW BGW BWG
4. WGB GWB WBG BGW GBW BWG

These sequences were then tested to evaluate if one arrangement is better than another. From the tests conducted it was observed that the order of the six combinations does not make a difference. Infact, sequence 2 is the same as sequence 3 read backwards. The only requirement is to avoid the number of possiblities of a gray line occurring between two white lines. Extending this constraint further, the possiblity of WBW can also be avoided to eliminate the thinning of a black line by two white lines. It can be observed that sequence 1 and 4 have two such possibiities, while 2 and 3 have only one occurance of WBW.

The three level coding scheme has a definite advantage over the binary coding scheme with step like objects as shown in figure 3.19. When changes in depth distort the projected pattern, it is still possible to trace the lines based on the known sequence. From figure 3.19, it can be observed that knowledge of the sequence helps to trace the lines, which are shifted and thereby false matching can be avoided.

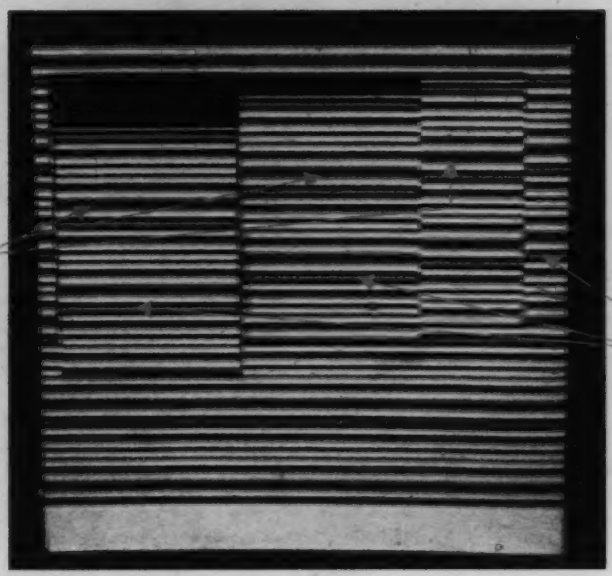


Figure 3.19: Tracing broken lines with three-level pattern

However the problem of occlusion still persists. Along the same lines of three level codes, a number of slides were also created with 8 levels of gray ranging from 0-255. With 8 levels the number of possible combinations is higher than with three levels. However the lighting of the projector is not uniform over the entire range of the image, the frame grabber tends to identify 256 different intensities rather than the expected eight. The results from the three level codes also indicate that the captured image has the intensity range spread out over the 256 levels. This limits the use of multilevel (greater than 3) codes in the present system.

3.4 Colour Coding Scheme

This scheme uses more levels of intensities for labelling the surface. In general, due to different reflection properties of the object surface, the colour of the stripes recorded by the camera is different from that of the stripes projected by the light source. For example a green light stripe projected by the light source may be observed as blue after reflection. This problem is of great significance when we consider subjects with different tones of skin pigmentation. It is therefore difficult to solve the lighting to

image correspondence problem in many practical applications. However the colour observed by the two cameras in our system will be the same, even though this observed colour might not be exactly the same as the colour projected by the light source. A major advantage of using stereo vision together with colour stripes lighting is that, the more difficult correspondence problem between the light source and the image can be replaced by an easier image to image stereo correspondence.

Figure 3.20 shows a slide created in colour. With colour codes more combinations are possible. With more sequences, it is possible to reduce the repetition of a sequence and thereby eliminate the possibility of erroneous extraction of features and avoid false stereo matching. However, the use of colour-codes requires a colour camera system and a suitable frame grabber.

Figure 3.20: Colour coded slide

3.5 Analysis of the Results

From the experiments conducted with the different types of coding schemes, it is evident that three level and colour coding schemes are two promising techniques that can

be implemented to measure the surface topography of the human back. However the present camera-projector system precludes the use of colour coding schemes. This restricts the scope of this research, to the use of three-level codes. As discussed section, the binary coding scheme is not very successful with objects that have steep changes in depth, and therefore it is not possible to extract the features (lines) with certainty. Secondly the use of two level codes restricts the maximum number of lines (codes) that can be projected onto the surface. This would reduce the number of features that can be matched between the stereo images and thereby reduce the resolution of the measurement. With the three-level coding schemes it is possible to have lines, which are 4 pixels apart (vertical direction) and thereby increase the resolution of the feature extraction to 4 pixels. With the present camera-projection system it is possible to project lines of 4-pixel width (on the slide) onto a surface (such that it covers 4mm on the surface) and the width of lines in the captured image is 3 pixels. If all the captured pixels are matched in both the images, a matching resolution of 1.3 mm in the vertical direction can be achieved.

3.6 Justification for Selecting the Three-level Coding Scheme.

From the results of the above experiments, the three level code was found to be the appropriate coding scheme for the present projector-camera system. The sequence created with six combinations of three lines (white, gray and black) has a spatial frequency of 7.2cm (72 pixels, 1 pixel =1 mm). This covers 7.2cm on the object surface. Therefore steep rises on the surface would not affect the feature extraction. Each line in the sequence is uniquely defined by the subcode in which it is present.

As discussed in section 3.3, among the different sequences that are possible with the six combinations, only four can be used to reduce the possibilities of having WBW, and WBW occurring. Among the four combinations that satisfy these constraints the following code was chosen:

WBG WGB GWB GBW BGW BWG

This sequence has the combination WGW occurring when the sequence is repeated,

WBG WGB GWB GBW BGW BWG WBG WGB GWB GBW BGW BWG.

However during template matching this can be solved. The 640 x 480 pattern created using the above sequence is shown in figure 3.21.

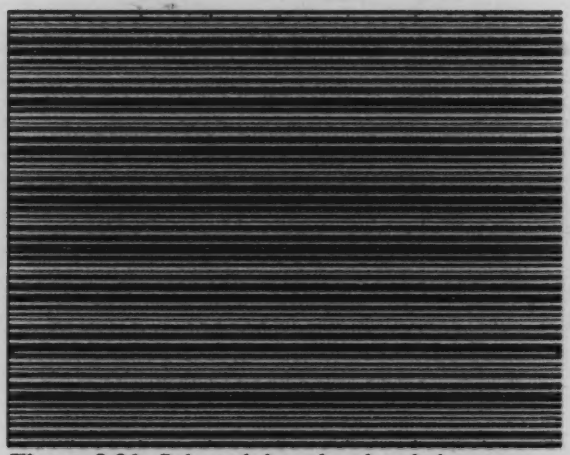


Figure 3.21: Selected three level coded pattern

This pattern has the sequence: WBG WGB GWB GBW BGW BWG

Where W(white)= wwww, B(Black)= bbbb, G(Gray)= gggg, w = 1 pixel of intensity level 255, b= 1 pixel of level 1 and g = 1 pixel of level 127.

This pattern can be projected onto the surface, the features can be extracted, correspondence matching can be performed. The image processing steps involved in matching the stereo images are discussed in detail in the following chapter.

4 Image Processing and 3D Surface Reconstruction

After the coding pattern has been selected, it is projected onto the surface to be measured. The pattern is deformed depending upon the surface topography. The stereo cameras capture this image. Image processing techniques such as image rectification and image correlation extract the deformed pattern from the stereo images. The stereo images are then matched for the extracted features by applying a stereo correspondence algorithm and the epipolar constraint. The image processing techniques that have been implemented are discussed in detail in the following sections.

4.1 Image Rectification

Rectification is a procedure for transforming the image into a plane parallel to the object plane. The rectification of a tilted or oblique image taken from a camera in space transforms the image into an equivalent orthogonal image taken from the same camera. Rectification can be produced in four basic ways: (1) graphically, (2) analytically (3) opto-mechanically and (4) electro-optically.

The plane of the image, which contains images of object points, is a two dimensional representation of a three dimensional object space. A three dimensional, right-handed Cartesian coordinate system X, Y, Z is used as reference for the object space (figure 4.1).

From the camera calibration the elements of interior orientation of the camera are known. These elements are the principal distance, principal point location and lens distortions. Exterior orientation of the image plane defines its position and orientation in object space. The position of the image plane is defined by the object space coordinates of the perspective center X_L, Y_L, Z_L. The orientation, which describes the attitude of the camera, refers to the spatial relationship between the object coordinate system (X, Y, Z) and the image coordinate system (x, y, z). The relationship between the image and the object coordinate systems is expressed by a 3 X 3 orthogonal matrix designated by M.

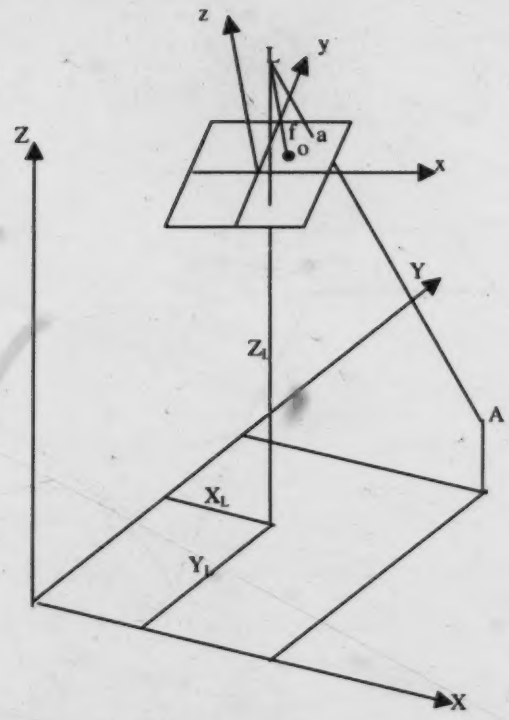


Figure 4.1: The object coordinate system.

The nine elements of M are functions of the orientation angles ω, ϕ, κ , which are the rotation angles of the camera with respect to X, Y, Z axes. The orientation matrix M is computed to transform from the object space system X, Y, Z to the image space system x, y, z.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = M \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad \text{(Eq. 4.1)}$$

In the present system, the two stereo images can therefore be rectified by a plane to plane resection as illustrated in figure 4.2 where,

$$\begin{bmatrix} x_n \\ y_n \\ -c \end{bmatrix} = M \begin{bmatrix} x_o \\ y_o \\ -f \end{bmatrix} \qquad \text{(Eq 4.2)}$$

Where x_o, y_o, are the x, y coordinates of the points after image capture and f is the focal length of the camera, and x_n, y_n are the rectified coordinates and c is the new focal length. The value of c is fixed for both the cameras so that they can be matched by using the epipolar constraint.

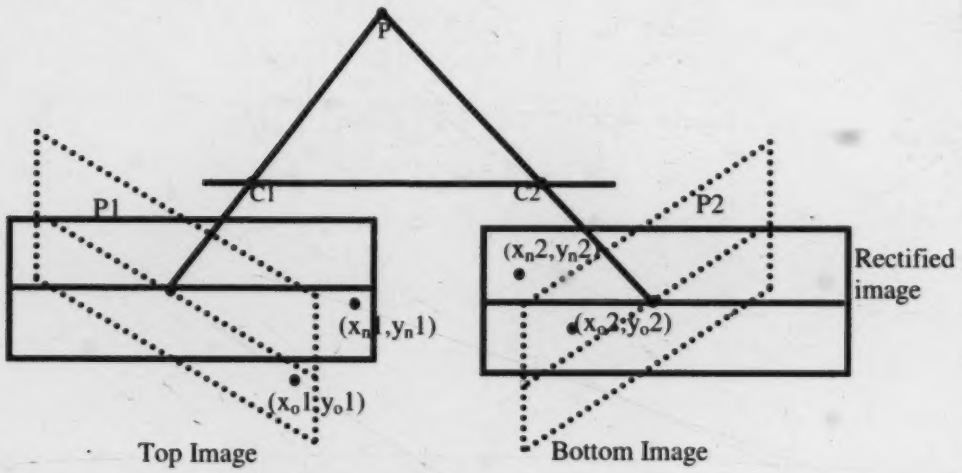


Figure 4.2: Rectification of two images

The cameras are calibrated and therefore the parameters of both stereo cameras are known. Based on the exterior orientation of the stereo pair the rotation matrix M, is computed.

$$M = R_\omega R_\phi R_\kappa \qquad \text{(Eq 4.3)}$$

where,

$$R_\omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & \sin\omega \\ 0 & -\sin\omega & \cos\omega \end{bmatrix} \qquad \text{(Eq. 4.4)}$$

$$R_\phi = \begin{bmatrix} \cos\phi & 0 & -\sin\phi \\ 0 & 1 & 0 \\ \sin\phi & 0 & \cos\phi \end{bmatrix} \qquad \text{(Eq. 4.5)}$$

$$R_\kappa = \begin{bmatrix} \cos\kappa & \sin\kappa & 0 \\ -\sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{(Eq. 4.6)}$$

Knowing ω,ϕ,κ, and f for both the cameras, the two images can be rectified. Eq 4.2 can also be written as

$$\begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} = M^{-1} \begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix} \qquad \text{(Eq. 4.7)}$$

Eq 4.7 was implemented to rectify the images so that the new rectified image consisted of points with gray level values known from the captured image.

4.1.1 Validation of the Rectification Procedure

The rectification approach described above was tested on a known set of points. The accuracy of the rectification program was found by rectifying a set of points with known coordinates. The rectified points were then compared to the known values to determine the accuracy. Table 4.1 shows the coordinates of a set of selected points that were tilted 20^o about the Y-axis, with known rectified points. Table 4.2 shows the coordinates of the same points when rectified. The points were selected in such a way that they occur on both sides of the origin (center of the image). The program then takes the data in Table 4.1 as the input and computes the rectified points. Table 4.3 shows the output of program. By comparing table 4.2 and table 4.3 it was observed that the error in computing the rectified points is less than 1 pixel (round off error). Another trial was conducted with the data in table 4.2 as input and the angle of rotation was complemented, the output of the rectification program was compared with the data in table 4.1. The data

in table 4.2 was taken as the input data and the rectification subroutine generates the output, which were compared to the data in table 4.1 and error in the computation of the rectified points was less than 1 pixel (round off error). To test the program on the captured image an image of a tilted plane was rectified. Figure 4.3 is an image of a tilted plane. The rectified image of 4.3 is shown in figure 4.4.

The accuracy of the rectification process is dependent upon the calibration data. The rectification program requires the accurate computation of the focal length and the angle of tilt of the cameras. The rectification program also corrects for the aspect ratio of the cameras. For the Sanyo cameras, this factor was determined from the average ratio of the principal distance in the x-direction to that in the y direction for a number of calibration runs. The ratio was found to be 0.9855:1(x:y).

Point number	X-coordinate	Y-coordinate
11	.0000	100.0000
13	100.0000	100.0000
21	.0000	0.0000
23	100.0000	0 .0000
31	.0000	-100.0000
33	100.0000	-100.0000
111	.0000	111.1111
121	.0000	0 .0000
131	.0000	-111.1111

Table 4.1: Test data used as input for rectification

Point number	X-coordinate	Y-coordinate
11	65.5146	127.7013
13	244.9513	168.6153
21	65.5146	0.0000
23	244.9513	0.0000
31	65.5146	-127.7013
33	244.9513	-168.6153
111	65.5146	141.8904
121	65.5146	0.0000
131	65.5146	-141.8904

Table 4.2: Known rectified points of table 4.1

Point number	X-coordinate	Y-coordinate
11	65	128
13	245	169
21	65	0
23	245	0
31	64	-128
33	245	-169
111	65	142
121	65	0
131	65	-142

Table 4.3: Computed rectified points of table 4.1

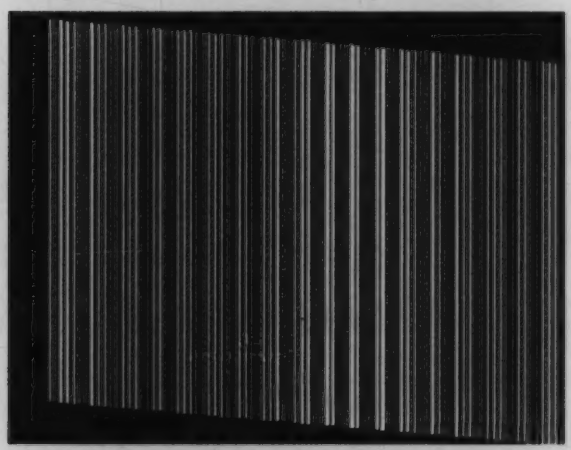


Figure 4.3: Tilted image of a plane

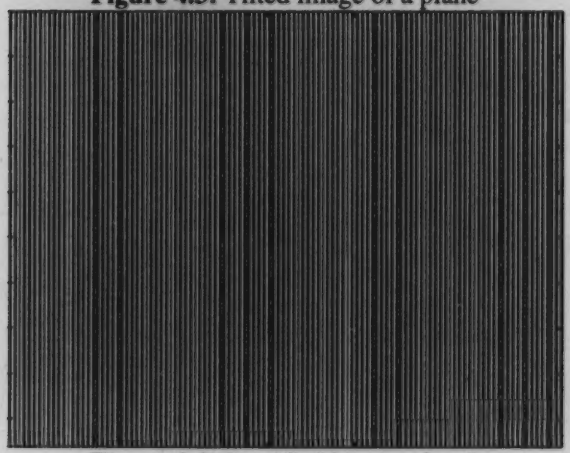


Figure 4.4: Rectified image of a plane

4.2 Feature Extraction

This section focuses on the implementation of the algorithm for automatic feature extraction. After the images are rectified, it is sufficient to confine the search of conjugate features along the same scan lines. However, for camera systems with low spatial resolution, the spatial frequency is higher and aperiodic. Therefore it is not possible to identify the coded pattern based on the spatial frequency directly since the pattern does not have the same thickness as the one projected onto the subject. The technique of template matching or correlation can be used to extract features (location of the codes in the present system) to solve this problem.

Template Matching

The pattern of projected light is coded and it is necessary to identify the code locations on the captured image. The template matching approach can be used to solve this problem. The problem of template matching can be stated as: given an image, g(x,y) of size M x M, and a given template, t(x,y) of size N x N (where N < M) , find the locations in g(x,y) whose neighborhoods closely resemble t(x,y) [1]. One of the significant steps in template matching is template design. The size of the template should be determined so that the required features can be extracted from the captured image. The following sections discuss this process in detail.

4.2.1 Template Design

As discussed in Chapter 3, the coded pattern selected was

WBG WGB GWB GBW BGW BWG

Where B represents 4 pixels of black intensity, W represents 4 pixels of white and G stands for 4 pixels of gray. Previous experiments demonstrated that a 4mm code on the surface is scaled to 3 pixels in the captured image. A 72 pixel code that covers approximately 72mm on the surface of interest would be approximately 54 pixels wide

on the captured image. Therefore a template of size 72 x 1 would not yield good correlation results.

An image of a plane surface (figure 4.5) was captured. A single column of the image was then sectioned and the intensity along that column was plotted. Figure 4.6 shows a column of image and figure 4.7 shows the intensity variations along a part of that column. From the captured image it can be observed that each line is 3 pixels wide. A template was created, where each line is 3 pixels wide. As there are 6 combinations and each combination has 3 lines (white black and gray) this template would have a size of 54 pixels (6 x 3 x3 pixels).

A template of size 54x1 pixels was tested by correlation. Figure 4.8(a) shows the plot of the 54-pixel template and figure 4.8(b) shows the plot of the intensity of the captured image along the column. The 54-pixel template was correlated over the captured image and it can be observed from figure 4.8(c) that the two plots align along the 54 pixels. Correlation was carried out for the entire column. Figure 4.8(d) shows the results of the correlation for a single column. The maximum normalized correlation coefficient is 0.95 at the peaks. The peaks indicate the points where the 54-pixel template matches (best correlation) the 54 pixels of the captured image. The number of peaks represents the number of times the template matches well with the captured image for that column of the image. In figure 4.8(c) there are 5 peaks which indicate that there are 5 points where the template matches well with the captured image. These locations were then compared to the actual points where the template matches with the captured image. It was found that the correlation results correspond with the manual matching process.

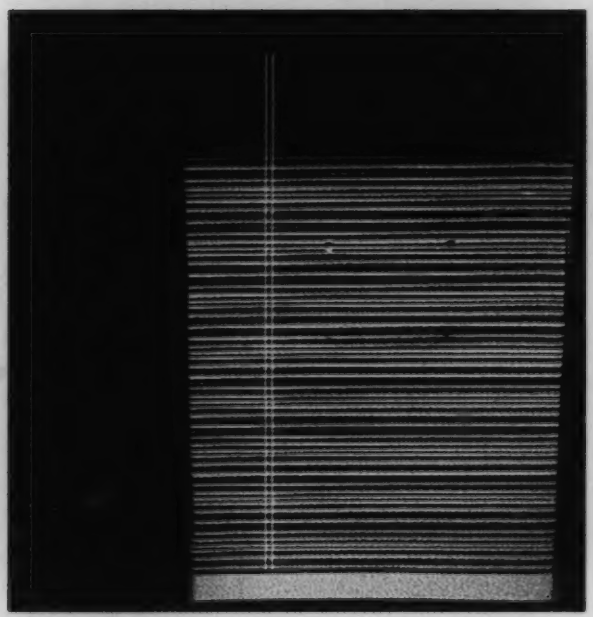


Figure 4.5: Captured image of a plane surface

Figure 4.6: Column of captured image

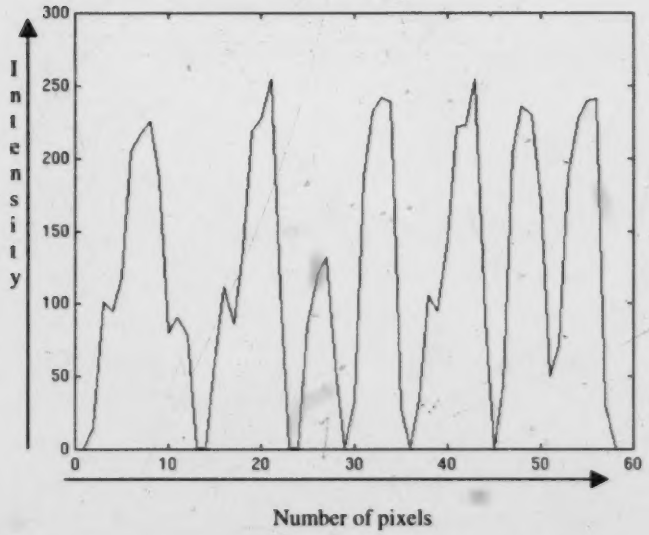


Figure 4.7: Plot of intensity vs Pixels along the column of the captured image

From the plots (Fig 4.8(a-d)) it was found that when the template size was 54 x 1 the template aligns well over the captured image and the correlation results validate this observation. (figure 4.8(a)). The 54-pixel template was therefore used in feature extraction.

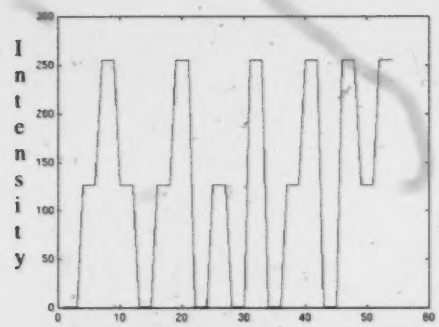


Figure 4.8(a): 54pixel template

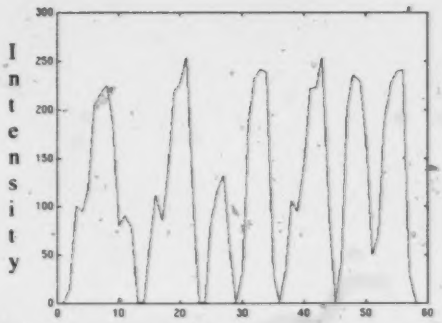


Figure 4.8(b): A plot of the intensity of the captured image along a column

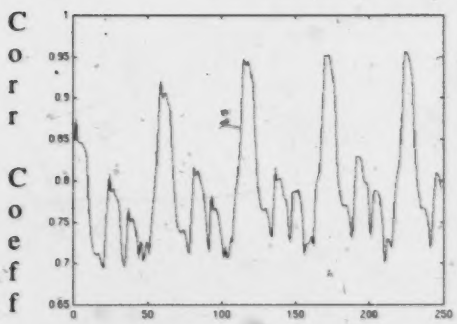


Figure 4.8(c): Correlation results

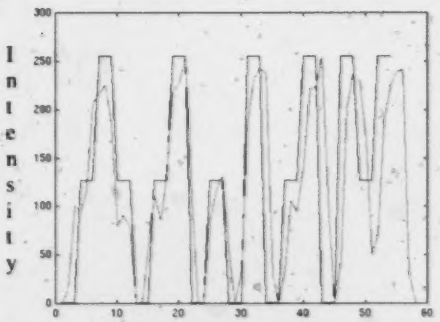


Figure 4.8(d): 54 pixel template overlaps well over the plot of the captured image

4.2.2 Implementation of Feature Extraction by Correlation

After the 640 x 480 image was captured, it was rectified and a column by column correlation was performed. That is the first column of 640 pixels was correlated with the designed template of 54 pixels. This located the points on the image that match well with the template and also the frequency of these matches was known for that column. This is illustrated in figure 4.9. (Locations of the best match are indicated by arrows)

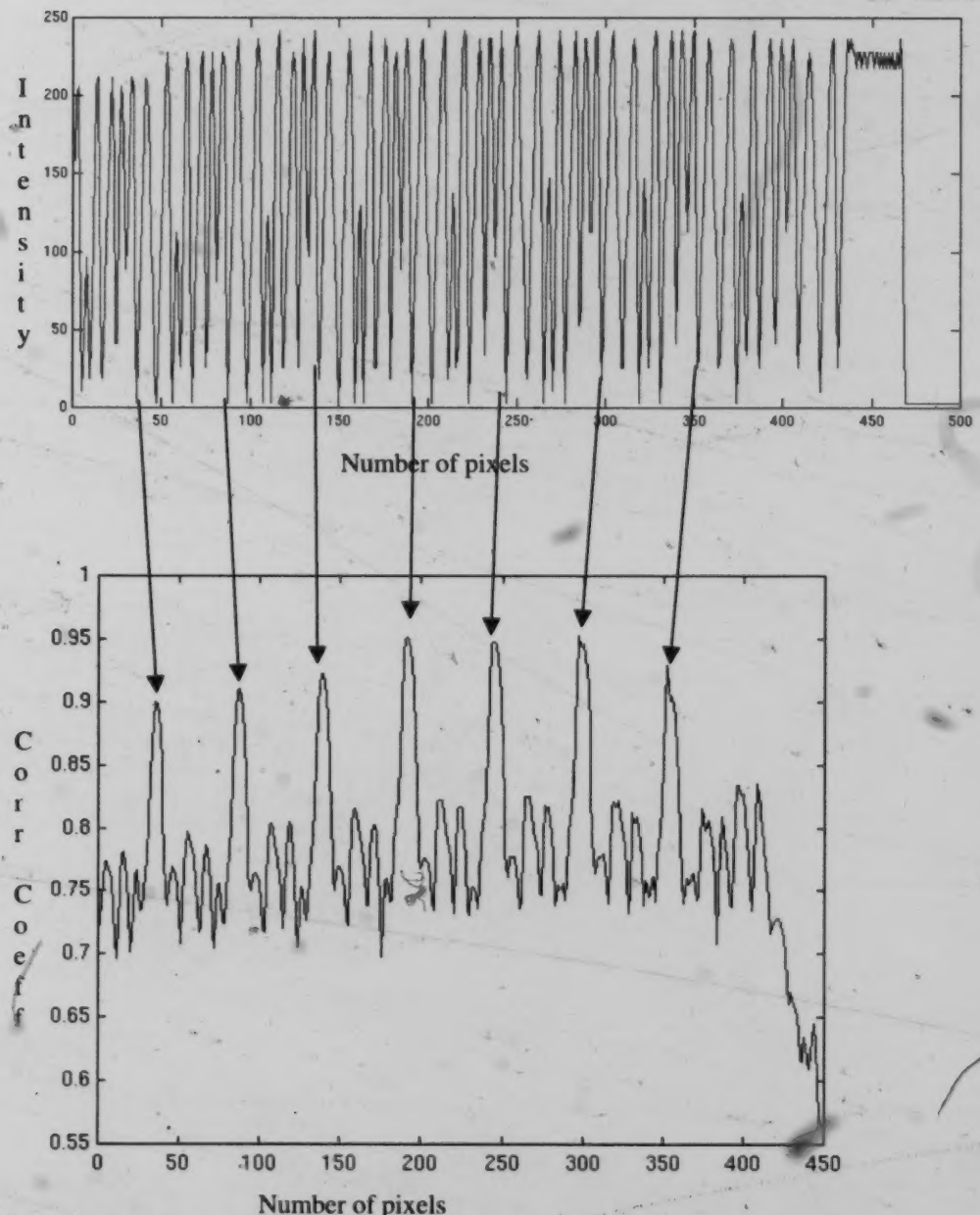


Figure 4.9: Plot of intensity along a column and the correlation results for that column

After these points were located in the image, the locations of the six combinations were determined. Each combination has 3 lines (B,W,G) and each line is 3 pixels (bbb,ggg,www) wide. Therefore a 9-pixel template can be used to locate these combinations. However with a 9-pixel template it is possible to extract the wrong feature. The six combinations to be extracted from the 54-pixel code are:

WBG WGB GWB GBW BGW BWG

1 2 3 4 5 6

To locate the 3rd combination, GWB, a 9-pixel template can be correlated over the entire sequence and the location of the best correlation can be determined, however there are two occurrences of GWB in the sequence. One is the 3rd combination in the sequence and the other is partly in the 5th and in the 6th combination as shown below.

WBG WGB **GWB** GBW B**GW B**WG

1 2 3 4 5 6

The correlation might therefore identify either of these combinations and thereby lead to false extraction. This is avoided by using a bigger template instead of a 9 pixel template so that the correct combination (the 3rd) can be picked up from the correlation results. The size of the new template can be from 12 to 54 pixels. Correlation is a computationally expensive process and the increase in template size would considerably increase the computational time required. To identify the location of the code GWB, a 12 pixel wide template can be used to locate GWBG. It can be observed that the distinction between the GWBG and GWBW is a single line of gray or white line each only 3 pixels wide on the captured image. Therefore any small amount of noise of 2-3 pixels can result in false extraction. A 15 pixel or 18 pixel template can be used to solve this problem. The 15 pixel template was not chosen as it would require that for every nine pixel code to be extracted the following 6 pixels for the template have to be derived from the projected code. This would increase the software overhead. If the template size is 18 pixels it covers two subcodes and with any change in the actual projected pattern the 18 pixel template can be determined automatically. Therefore an 18 pixel template was used to correlate over the 54-pixel region of the image to locate the subcodes. Since every set of

54 pixels must contain at least one set of WBG, WGB, BWG, BGW, GWB, GBW, the 18 pixel code covers two subcodes and is less prone to false extraction than the 12 or 15 pixel templates. To locate each 9 pixel subcode, a different 18 pixel template, specific to the subcode is required.

This process was then repeated for every column. The locations of all the subcodes would be known, for that particular column. The co-ordinates of all the nine-pixel wide codes were then stored in a file. The feature extraction process described above is illustrated in figures 4.10(a)-4.10(c). The 18 x 1 pixel mask (Figure 4.10(a)) is moved over the captured image (Figure 4.10(b)) and the location where the correlation is maximum is shown in figure 4.10(c).

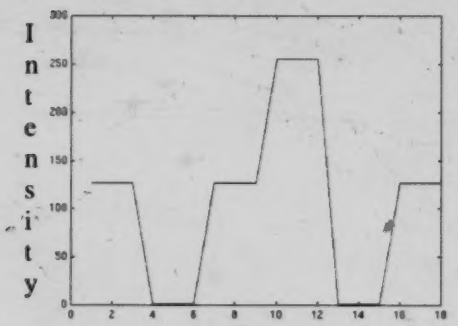


Figure 4.10(a): Plot of captured image

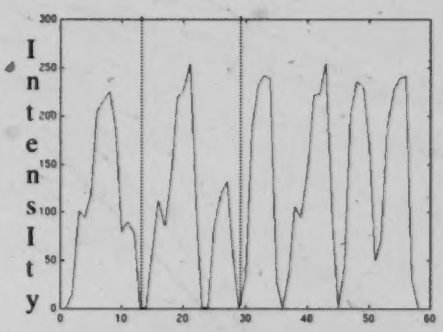


Figure 4.10(b): 18 pixel template

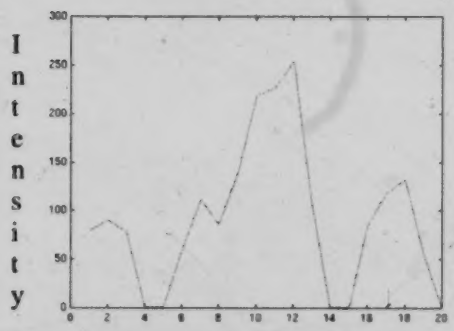


Figure 4.10(c): Plot of captured image along a column

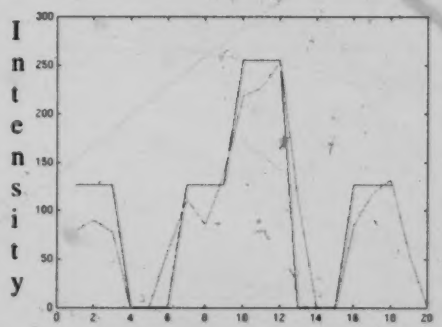


Figure 4.10(d): Plot of captured image over the 18 pixel template

Once these coordinates are known, it remains to identify the locations of the individual pixels in that subcode. A subcode, WBG, has approximately 3 white pixels, followed by 3 black pixels, which are followed by 3 gray pixels. The locations of the White, Black and Gray pixels have to be computed. One method is to identify the locations of the black pixels in the set of nine pixels and then identify the location of the white pixels. This can be done by searching for a pixel with the lowest intensity level in a set of nine pixels to identify the black pixels and then look for the nearest pixel with the highest intensity value, which will be a white pixel. This process would identify the location of all the black and white pixels. The location of gray pixels cannot be determined with certainty because the jitter effect of the frame-grabber makes it difficult to distinguish the gray pixels from the white and black pixels. This is illustrated in figure 4.11. The figure shows the B GWB GB section, of the 54-pixel sequence. The intensity values of the pixels are also shown. It can be observed from figure 4.11 that while the lowest and highest intensity level can be used to identify white and the black pixels, there are a few pixels which cannot be classified. In figure 4.11 pixels with intensity levels 41, 3 and 41 can be identified as black, based on the lowest intensity levels. Similarly pixels with intensity levels 212, 237, 237 can be identified as white pixels. If an attempt is made to locate the gray pixels based on elimination techniques then pixels with intensity levels 113 and 147 (figure 4.11) might be falsely identified as gray pixels and this would lead to incorrect feature extraction. In the present system it was therefore decided not to extract the gray pixels. The role of the gray pixels in the present system would be to give every subcode (comprised of B,W,G) a unique label. This helps to extract the white and black lines by correlation with ease and certainty. With high resolution projector camera system attempts can be made to identify the gray pixels and thereby further increase the resolution of the system.

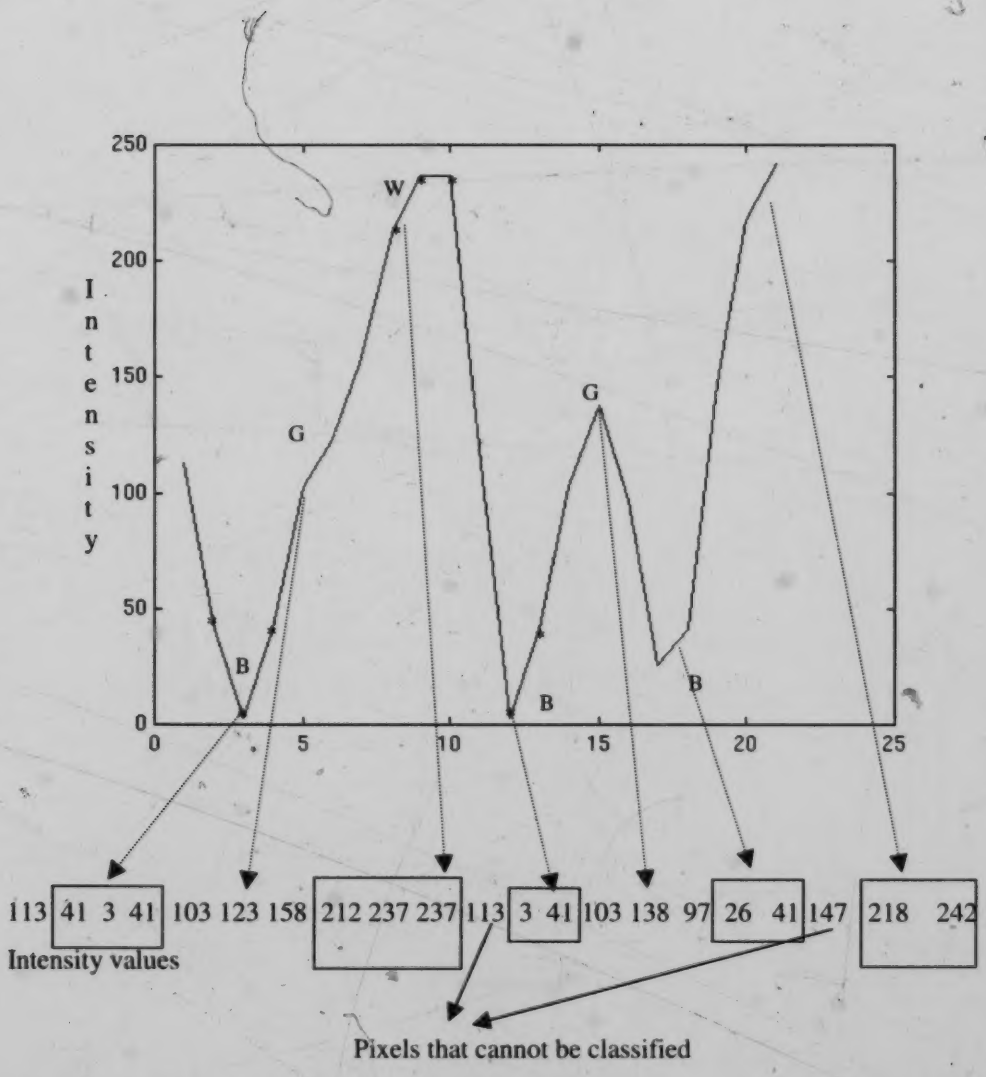


Figure 4.11: Plot of a subcode and the intensity values of that subcode in the captured image

4.2.3 Validation of Feature Extraction

To validate the feature extraction process by the template matching technique test images were generated on the computer. In these test images the locations of the gray, white and black pixels were known. The features were extracted from these images and were compared with the actual locations to evaluate the performance of the feature extraction module. Figure 4.12 shows a machine-generated image. The pixels with level 255 are white, with level 127 are gray and pixels with level 1 are black. Figure 4.13 shows the results of the feature extraction process. The generated image has no noise and all the points can be distinctly identified as one of the three levels and therefore the extraction process was successful in locating the pixels without any error.

Figure 4.14 shows another computer-generated image. This image is more complex. The code changes every second column but the sequence of the codes remains the same. This simulates a surface which has changes in surface depth every 2-mm, which is a worst case condition and is not typical in a human back. The results obtained after feature extraction are shown in figure 4.15. There was no error in extracting the features extraction. Based on the results obtained from these images further tests were conducted with images of test objects.

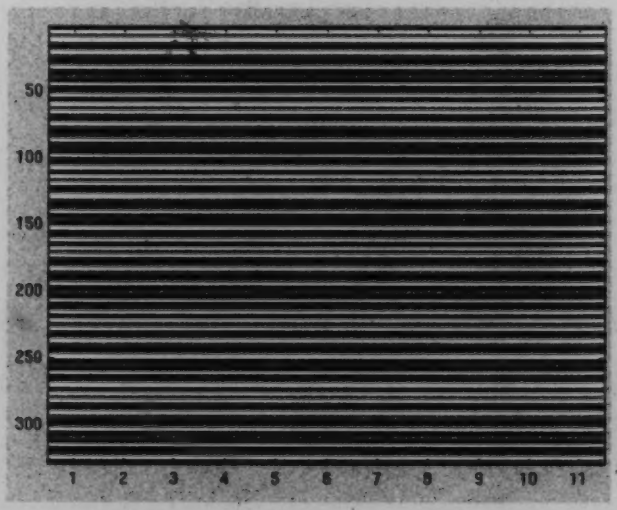


Figure 4.12: Computer generated test image

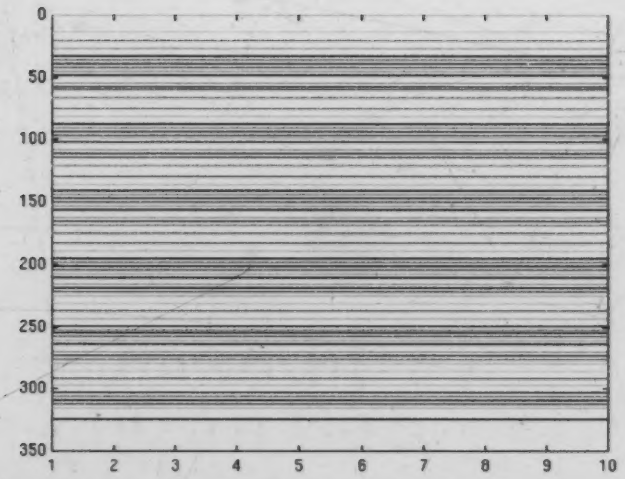


Figure 4.13: Results after feature extraction process

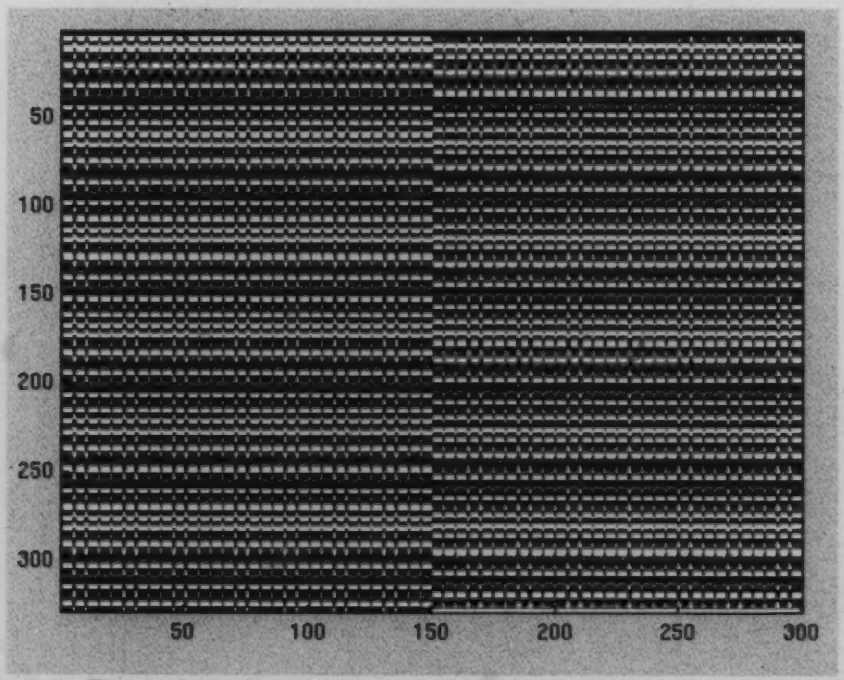


Figure 4.14: Computer generated test image

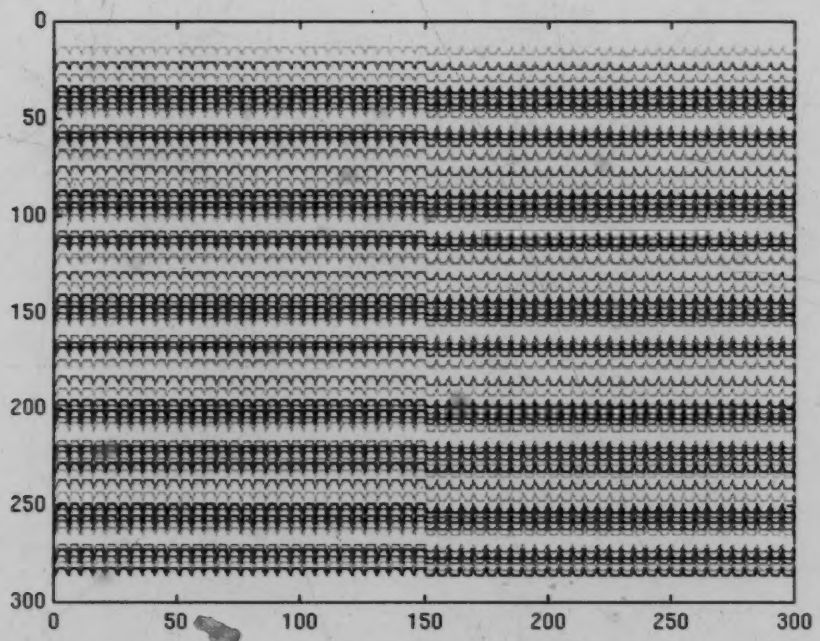


Figure 4.15: Results after feature extraction process

Figure 4.16 shows the captured stereo images of a ramp like object, and the results after feature extraction are shown in figure 4.17. From the extracted features it is possible to associate lines to the appropriate region of the coded sequence. The extracted features were compared with the actual location of the features on the image (by a manual process) the precision was one pixel. However when the lighting condtions are different the precision decreases. This is due to the jitter of the frame-grabber. This indicates that the correlation process is very sensitive to lighting conditions. This can be reduced by applying a filter to the extracted features. However, applying a filter at this stage of the processing has cumulative effects on image matching and triangulation steps to be carried out. The results of the feature extraction are promising and the stereo images can be matched for the extracted features. The maximum spacing between any two extracted features is 3 pixels.

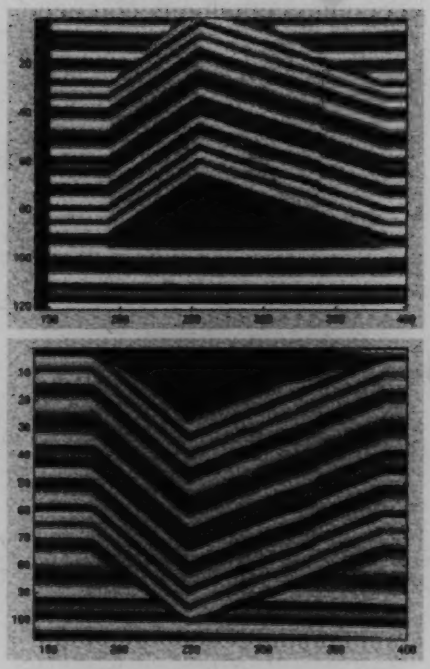


Figure 4.16: Stereo images of a ramp like object

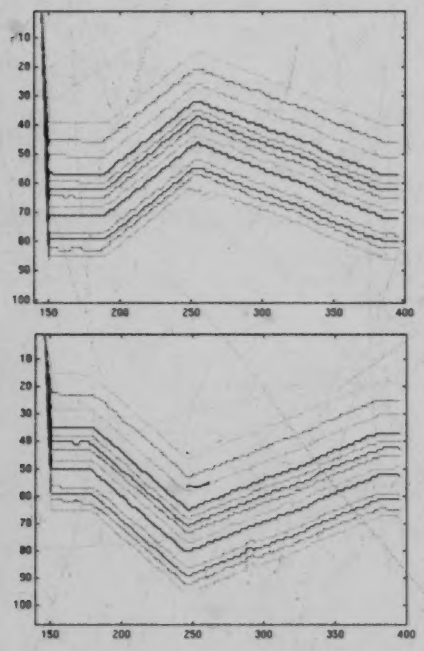


Figure 4.17: Results of the feature extraction process

4.3 Correspondence Matching

After the features are extracted from the stereo images, they have to be matched. This process is known as correspondence matching. The images are rectified and therefore, it is not required to search for every point in one image to locate the corresponding point in the other image. Such a two dimensional search is unnecessary because of a simple but powerful constraint: the epipolar constraint. It is well known that the epipolar constraint can considerably reduce the searching complexity [24], [45]. The stereo matching method used in this research is a variation of the intra/inter-scanline method proposed by Ohta and Kanade [32]. The major advantage of this method is that it is a global matching technique in which the correspondences are not only determined by local similarities, but also by global similarities. The implementation of the correspondence matching is discussed in the next section

4.3.1 Implementation of the Correspondence Matching

The objective of the correspondence matching module is to match the features extracted from the two stereo images. The features extracted from each image are the locations of a set of sub-codes for each column. The projected sequence is:

WBG WGB GWB GBW BGW BWG

Table 4.4 shows a sequence of the code for 4 columns of the top and bottom images.

After feature extraction the location of all the sub-codes is known. The location of the sub-code WGBGWB in the top image is at X4 (table 4.4) in column 1, while it is located at X7 in the bottom image in column 1. Similarly GWBGBW is located at X3 in column three and its conjugate in the other image is located at X6 in column 3. The feature extraction process generates column by column information of the location of each of the sub codes for each image. Since the images are rectified, by applying the epipolar constraint, it is possible to match the points in the two images. For the data in Table 4.4 the feature extraction process for the top image yields the location of the sub-code GWBGBW as X7, X4, X3, X7 for the columns 1 to 4 respectively. For the bottom camera image, the location of the sub-code GWBGBW would be X10, X7, X6, and X10. It is therefore possible to match X7, Y1 in the top image to X10, Y1 in the bottom image and X4, Y2 of the top image to X7, Y2 of the bottom image. This process is repeated for all the columns of the rectified stereo images. The first run of this process matches all the sub-codes in the stereo images. Once a sub-code, for e.g. GWBGBW is matched, it now remains to match the location of the individual pixels in that sub-code. Each sub-code has two sets of black pixels separated by two sets of white pixels and these are identified by the feature extraction process and the center black pixel (the one with the lowest gray level) is matched in the two images. Similarly the center white pixel is matched in the two images. The maximum separation between the white and black pixel is one set of gray pixels, which is 4 pixels of the projected pattern, this translates to 4mm along the vertical direction on the object surface.

X\Y	Y1	Y2	Y3	Y4
X1	W	W	G	W
X2	B	G	B	B
X3	G	B	G	G
X4	W	G	W	W
X5	G	W	B	G
X6	B	B	G	B
X7	G	G	B	G
X8	W	B	W	W
X9	B	W	B	B
X10	G	B	G	G
X11	B	G	W	B
X12	W	W	B	W
X13	B	B	W	B
X14	G	W	G	G
X15	W	G	W	W
X16	B	W	B	B
X17	W	B	G	W
X18	G	G	W	G

Top Image

X\Y	Y1	Y2	Y3	Y4
X1	B	W	B	B
X2	W	B	G	W
X3	G	G	W	G
X4	W	W	G	W
X5	B	G	B	B
X6	G	B	G	G
X7	W	G	W	W
X8	G	W	B	G
X9	B	B	G	B
X10	G	G	B	G
X11	W	B	W	W
X12	B	W	B	B
X13	G	B	G	G
X14	B	G	W	B
X15	W	W	B	W
X16	B	B	W	B
X17	G	W	G	G
X18	W	G	W	W

Bottom Image

Table 4.4: Sequence of codes as appears in the captured stereo images

For every column the complete code is present a finite number of times. For each 54-pixel code the above process is repeated. Therefore a column by column correspondence matching was performed and all the points that were extracted from the images were matched. The 3-D coordinates of the matched points were computed by triangulation.

4.4 Triangulation and 3D Surface Reconstruction

The 3-D content of a scene can be recreated using the "top" and "bottom' images of an object. In stereoscopy, the depth is estimated using triangulation. The estimation process requires the knowledge of the global position and orientation of each camera, the model of the cameras and the correspondence between all the feature points in both images. Given a single image, the three-dimensional location of any visible object point is restricted to the straight line that passes through the center of projection and the image of the object point. Consequently, with two independent images, we can locate the three dimensional position of any object point that is visible in both images to be at the intersection of two straight lines. The determination of the position of an object point in this fashion- is called triangulation.

The collinearity equations are based on the assumption that an object point, a perspective center of the camera, and an image point lie on a straight line. As a result, the collinearity equation is written as

$$\alpha = \kappa MA \qquad \text{(Eq 4.10)}$$

where α is a three dimensional image vector, κ is an unknown scale factor, and M is a 3 by 3 orthogonal transformation matrix. Parameter A is a vector from the perspective center to the object point. These equations are non-linear. By using Taylor's theorem, collinearity equations in linearized form are written for each matched point of one image as:

$$v_x = b11dX + b12dY + b13dZ + c_1 \qquad \text{(Eq 4.11)}$$
$$v_y = b21dX + b22dY + b23dZ + c_2 \qquad \text{(Eq 4.12)}$$

where

$$b_{11} = (x/q)m31 + (f/q)m_{11}$$
$$b_{12} = (x/q)m32 + (f/q)m_{12}$$
$$b_{13} = (x/q)m33 + (f/q)m_{13}$$
$$b_{21} = (y/q)m31 + (f/q)m_{21}$$
$$b_{22} = (y/q)m32 + (f/q)m_{22}$$
$$b_{23} = (y/q)m33 + (f/q)m_{23}$$
$$c_1 = (qx+rf)/q$$
$$c_2 = (qy+sf)/q$$

and

$$r = m_{11}(X-X_o) + m_{12}(Y-Y_o) + m_{13}(Z-Z_0)$$
$$s = m_{21}(X-X_o) + m_{22}(Y-Y_o) + m_{23}(Z-Z_0)$$
$$q = m_{31}(X-X_o) + m_{32}(Y-Y_o) + m_{33}(Z-Z_0)$$

m_{ij} are the elements of the orthogonal transformation matrix, x and y are the image coordinates of the point after correction for the camera distortions, X_o, Y_o,and Z_o are the coordinates of perspective center of the camera, and f is the focal length of the camera.

After a pair of matched points was determined from the top and bottom camera images, a set of four equations based on Eqs 4.11 and 4.12 was obtained. The three unknowns Dx, Dy, and Dz were computed using the least squares method [45]. The camera calibration data and the set of matched points between the two images were the input to the program and the 3-D co-ordinates were the output. Figure 4.18 shows the 3-D points of the ramp surface computed by triangulation.

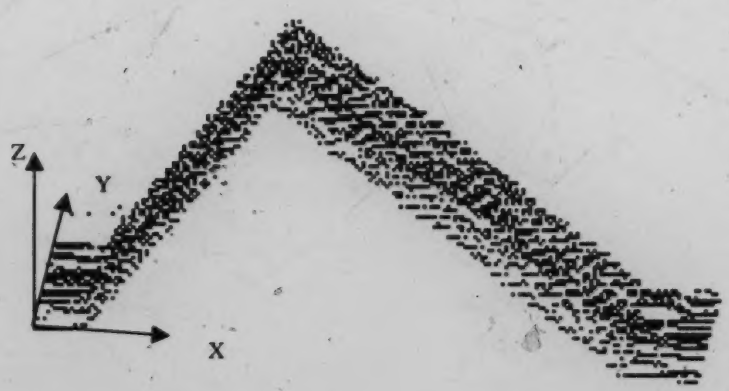


Figure: 4.18: 3-D points of the triangular surface computed by Triangulation.

Figure 4.19 shows the various steps involved in the surface reconstruction. The techniques that have been described were tested with known objects. The results of the experiments are discussed in the following chapter.

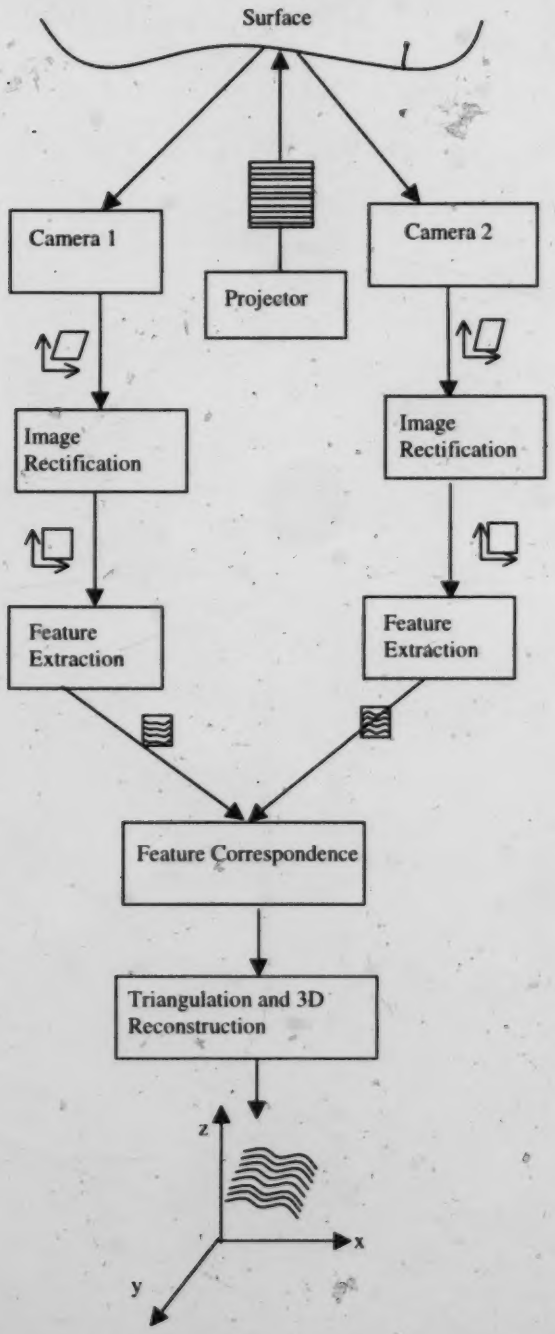


Figure 4.19: Various steps involved in 3D surface reconstruction

5 System Validation and Testing

To determine the accuracy and the precision of the system, test objects and surfaces were selected. Tests were conducted with these objects to determine how closely the objectives were met by the chosen methods. These tests established the accuracy, reliability, speed of process, and the level of automation. The quality and the accuracy of different stages of surface reconstruction of the back are discussed in separate sections in this chapter. The results obtained and the limitations of the system are also discussed.

5.1 Reconstruction of Test Objects

Figure 5.1 shows the stereo images of a test object with a curved surface. The height of the object is 85mm and the width is 288mm. This object was chosen as a test model. It has a uniformly curved surface with no sudden changes in depth, which is typical of a normal back surface. The various steps involved in computing the surface topography of the object are carried out. The following steps describe the three-dimensional surface reconstruction of this object.

1. Camera Calibration:

The stereo cameras are calibrated by capturing the image of the control frame, as described in chapter 3. The calibration results are therefore very critical for the accuracy of the system. A small difference in calibration results can significantly affect the computed depth. The manual digitization of the points on the control frame changes the calibration results from one trial to another.

2. Image acquisition:

The object is placed in front of the cameras and the known pattern of light is projected on its surface. The stereo images of the object are taken. The long axis of the camera is aligned with the long axis of the object so the cameras were rotated about 90 degrees about the Z-axis. Figure 5.1 shows the stereo images of the object taken by the top and bottom cameras (rotated by 90 degrees).

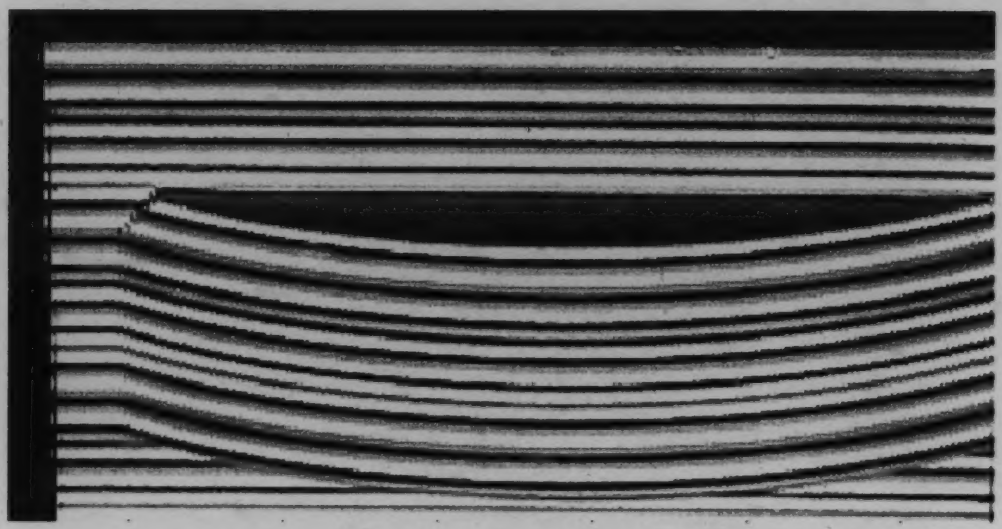


Figure 5.1(a): Top Camera view of the object surface

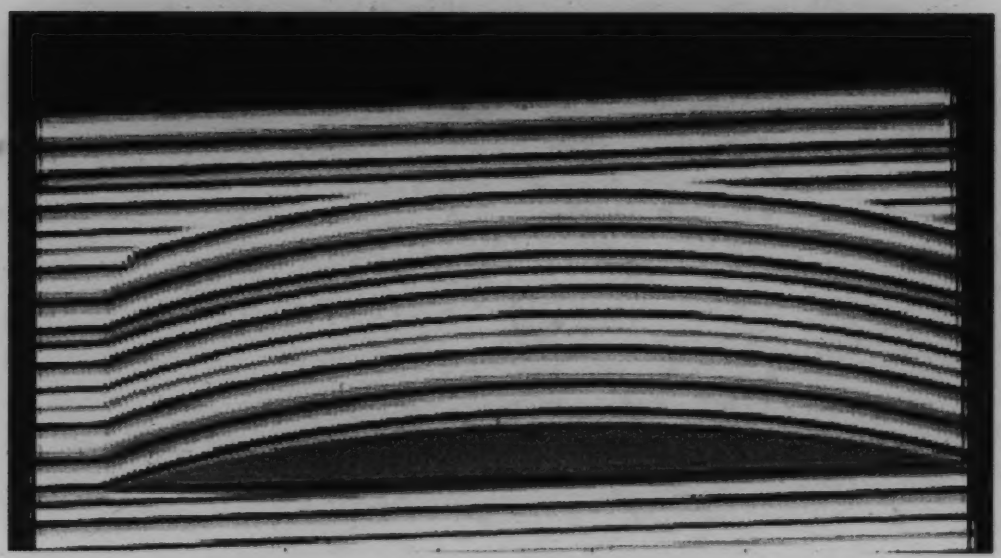


Figure 5.1(b): Bottom Camera view of the object surface

3. Determining the Region of interest:

After the stereo images are captured, the images are rectified so that the epipolar constraint can be applied. The region of interest in each image is then determined. The image processing steps discussed in the previous chapter are applied only to the region of interest. The region of interest is determined in two steps. The first step is to identify the 'x' co-ordinate where the region of interest starts and then locate the end of the region of interest. In our application the region of interest can be identified as the region, which has the subject towards the center and is surrounded by a uniform dark region. The program scans the image, starting from the left most column and looks for the first occurrence of a white region which follows a uniform dark area. The first such hit is taken as the starting point of the region interest. To locate the point where the region of interest ends, a similar scanning is done along the column and when a continuous region of dark area is located the starting point of the dark region is taken as the end of the region of interest. This process is repeated along the rows to determine the starting and ending points of the region of interest. After determining the region of interest, the feature extraction process is applied.

4 Feature Extraction:

The feature extraction process is done by correlating the template over the column of the image. For every column, the location of the template is determined and the required features in that template are extracted. The extracted features are shown in figure 5.2. The feature extraction process is then carried out for the other image and the same features are extracted in the second image as in the first. Figure 5.2(a) shows the features extracted from the top camera's image and figure 5.2(b) shows the features extracted from the bottom camera's image. The extracted features from the 54 pixel code are as shown below.

BGWGBGWBGBWBGWBWGW

The 12 features extracted from each column are shown in separate colours.

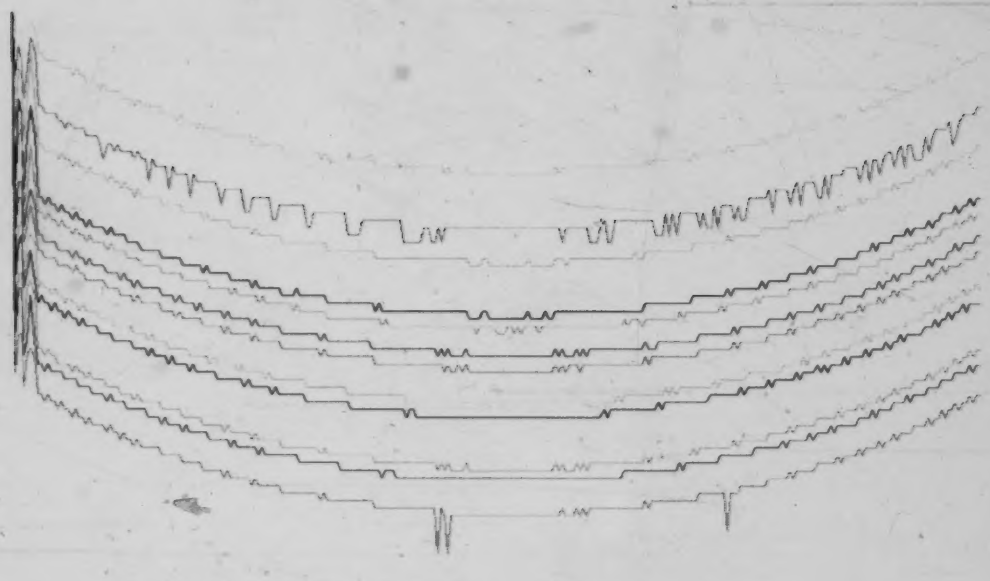


Figure 5.2(a): Results of feature extraction of the top image

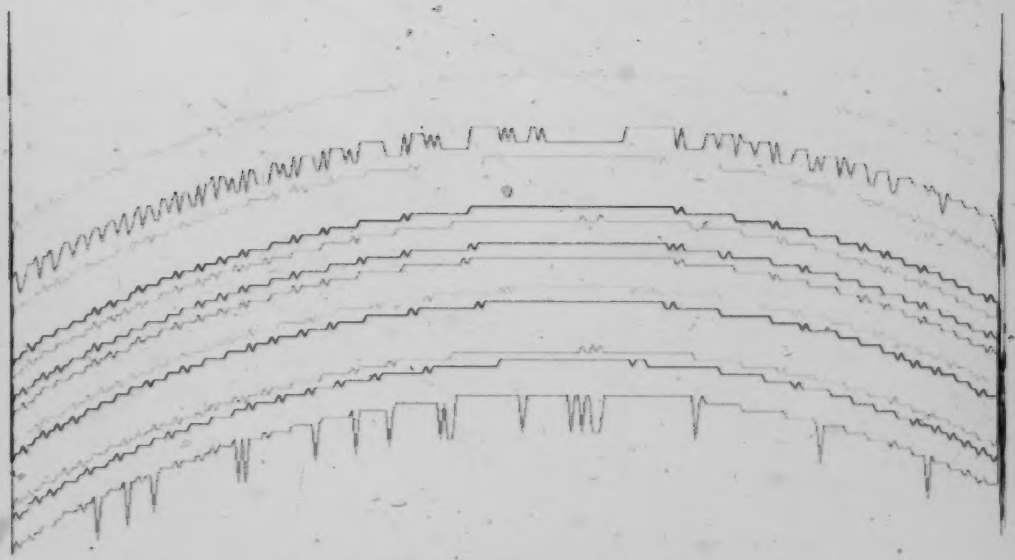


Figure 5.2(b): Results of feature extraction of the bottom image

5 Correspondence Matching.

After the features are extracted from the rectified images, the next step is to match a feature in one image to that in the other in the other image. Applying the epipolar constraint, a column by column matching is performed for all features extracted.

6 Triangulation

The matched sets of points are then used to compute the 3-D data. The program (SPACO) [1] was used to generate the 3-D points from the matched pairs.

The computed points are then plotted to view the surface of the object captured and measurements were performed. The figure 5.3 shows the 3D- surface of the object. 5.3(a) and 5.3(b) are different views of the surface. Table 5.1 shows the results obtained and the error in measurements.

Parameter	Known	Calculated	Error
Height	85mm	83mm	2mm
Length	288mm	283mm	5mm
Depth	53mm	52mm	3mm
Radius	224mm	219mm	5mm

Table 5.1: Results for the curved surface.

The measurement was repeated thrice and the precision of the system was determined. With the object at three different heights, images were captured and measurements were performed. The results are shown in table 5.2

Parameter	Trial 1	Trial 2	Trial 3
Height	83mm	83mm	84mm
Length	283mm	283mm	283mm
Depth	52mm	53mm	52mm
Radius	219mm	219mm	219mm

Table 5.2: Results from the 3 trials conducted

However as the object is not affected by movement during these measurements and there are not many factors that change from one trial to another in fixed objects the results are accurate and precise to 1mm. The precision has yet to be measured on human subjects where breathing effects and patient posture can challenge the precision of the system.

Figure 5.3: 3D view of the surface from two angles

However, surfaces of the backs of patients with scoliosis can have steep or ramp like surfaces or sudden changes in depth. To evaluate the performance of the system under such conditions a ramp-like object was tested. As discussed earlier, the stereo images of the object are captured (figure 5.4) and images are rectified and feature extraction process is carried out. The results of the feature extraction process are shown in figures 5.5(a) and 5.5(b). The correspondence matching is then performed and the 3-D surface generated. Figures 5.6(a) and 5.6(b) show the 3-D surface of the ramp like object. The measurements were made from the 3-D points computed. The results are shown in Table 5.3.

Parameter	Known	Calculated	Error
Height	75mm	74mm	1mm
Length	200mm	205mm	5mm
Depth at center	75mm	73mm	2mm

Table 5.3: Results for ramp surface

In extreme cases of scoliosis the surface can have step like features. Tests were performed to account for such cases. A step like object as shown in figure 5.7. The test object has three steps, the first step is 10mm, the second is 3mm and the third step is 10mm. This is to test the performance of the system when there are sudden changes in the surface. Another purpose of this test is to determine the performance of the feature extraction process when step like surfaces are encountered. In such objects the steps shift the projected lines and in extreme cases the lines may be discontinuous. This is due to the fact that the cameras integrate the pixels and the captured image may have disjoint lines in the vicinity of the steps. As before the 3-D points are computed and the surface is plotted. Figure 5.8 shows the result of feature extraction. Figure 5.9 shows the surface that was generated. Measurements were carried out and were compared with the known values. The results are tabulated in table 5.4.

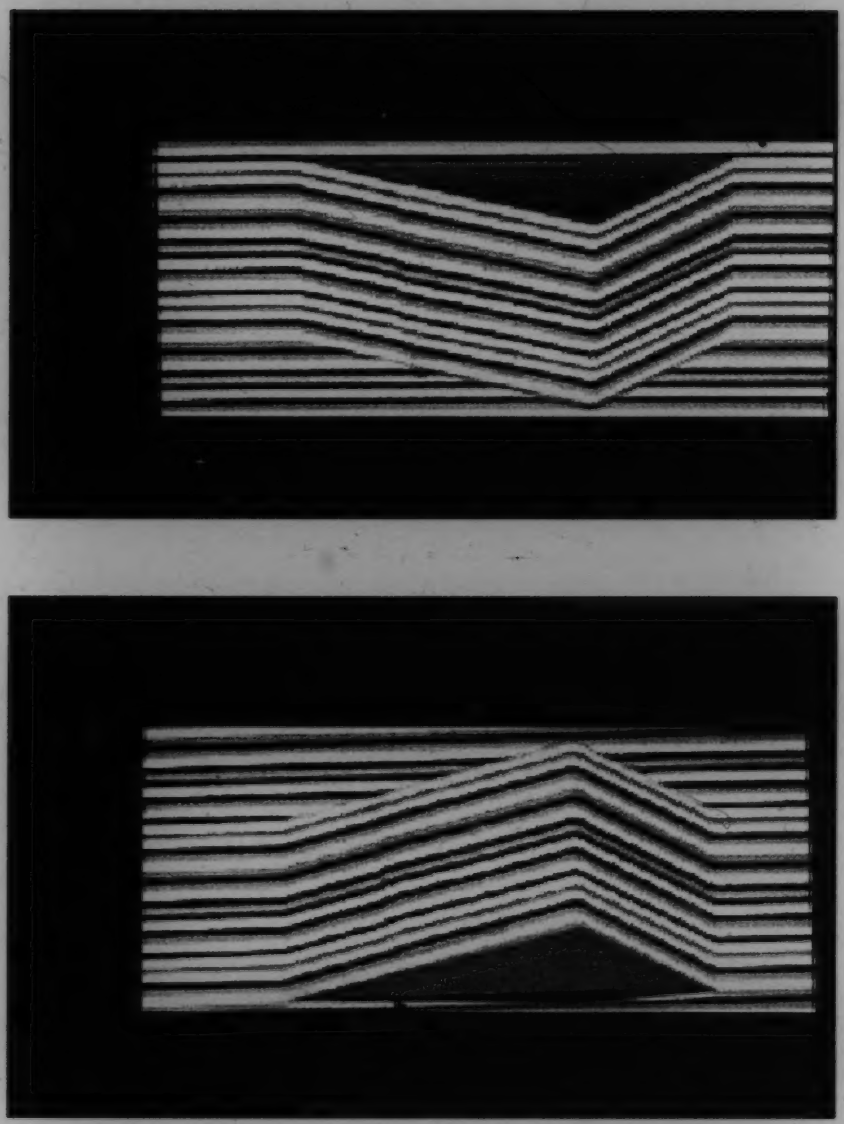


Figure 5.4: Captured stereo images of the triangular object

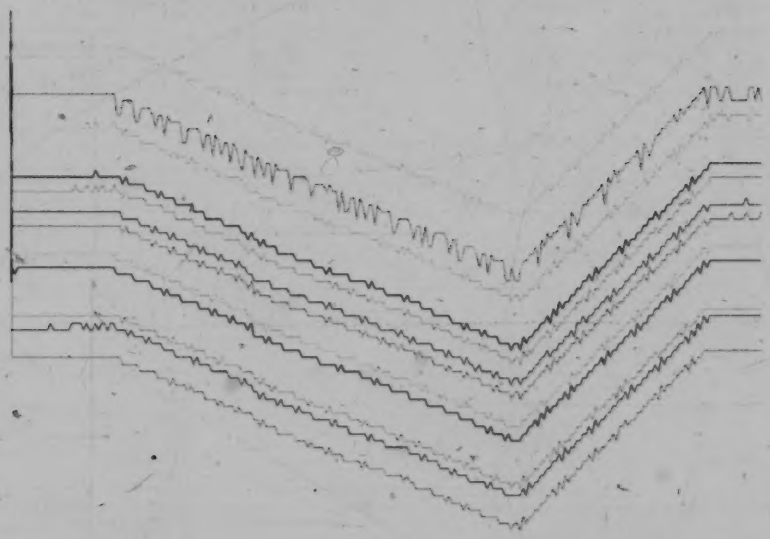


Figure 5.5(a): Results of feature extraction from the top image

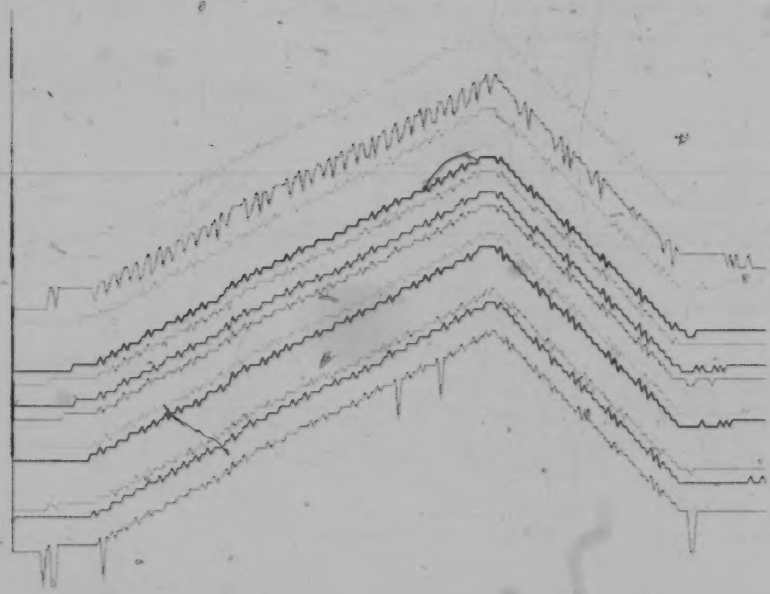


Figure 5.5(b): Results of feature extraction from the bottom image

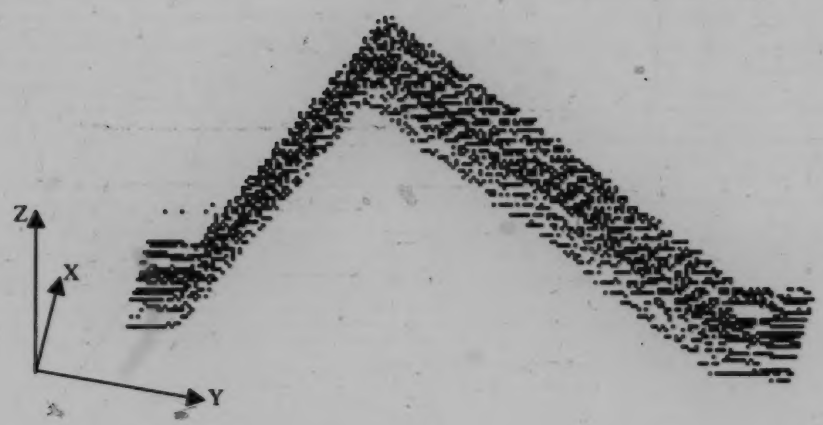


Figure 5.6(a): 3-D View of the surface

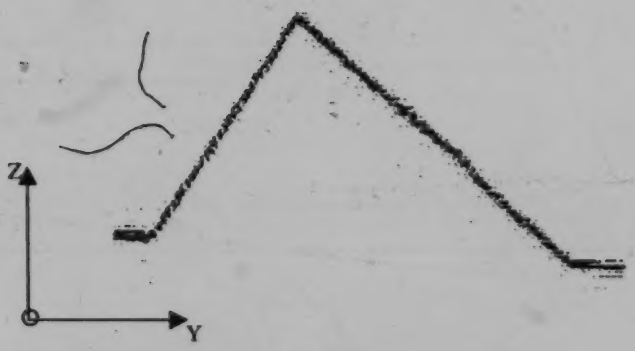


Figure 5.6(b): 3D view of the ramp surface

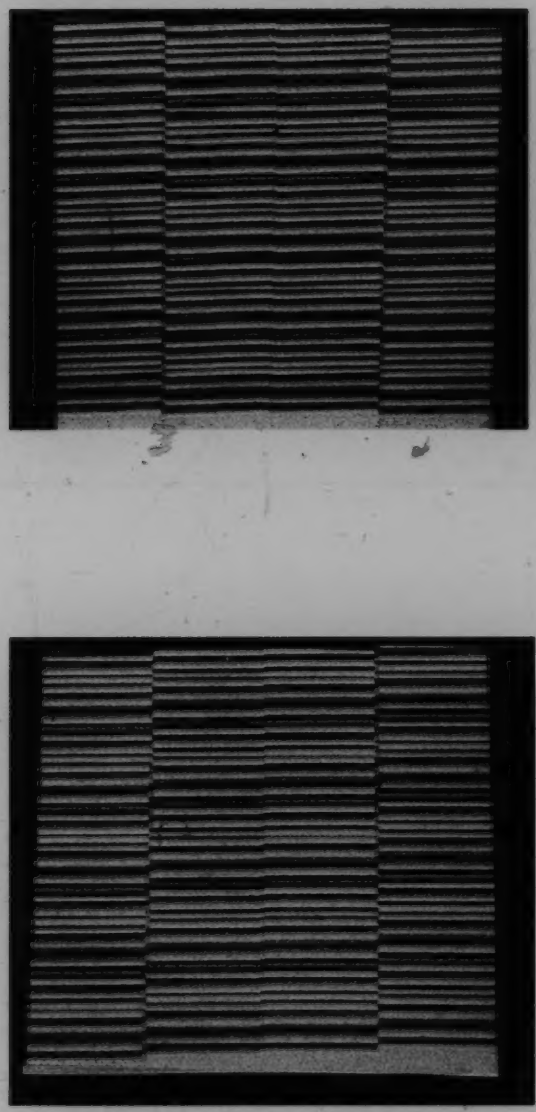


Figure 5.7: Stereo images of the step like object.

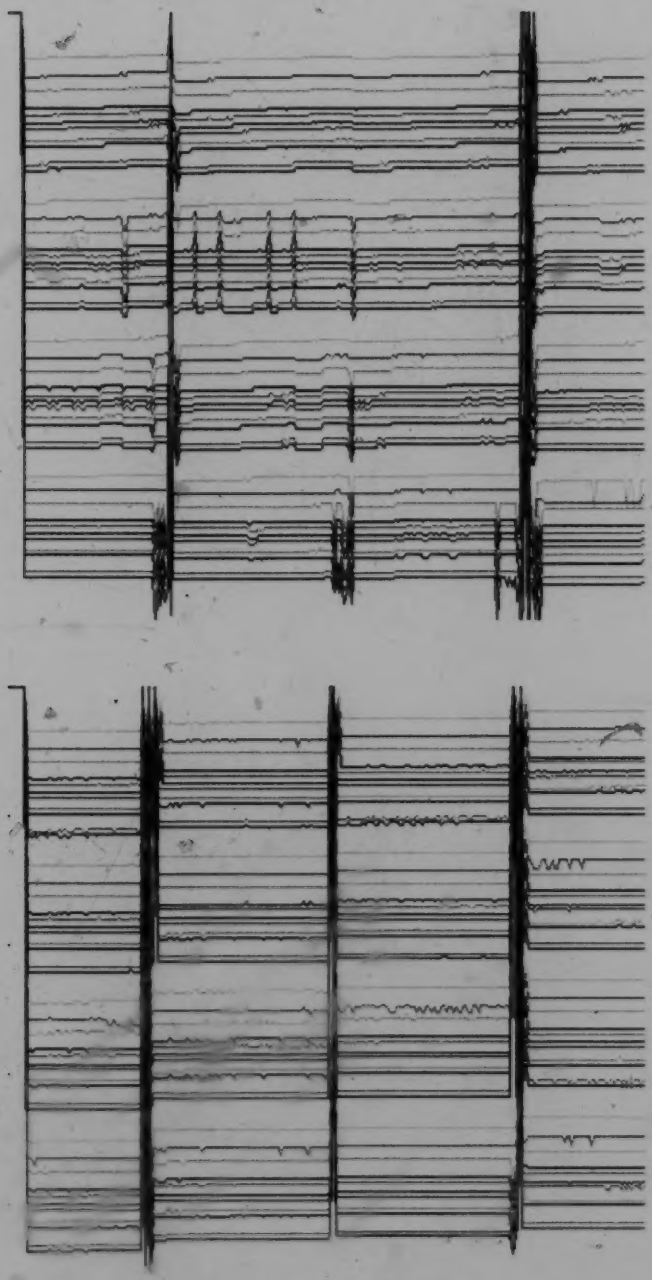


Figure 5.9: Results of feature extraction for step object

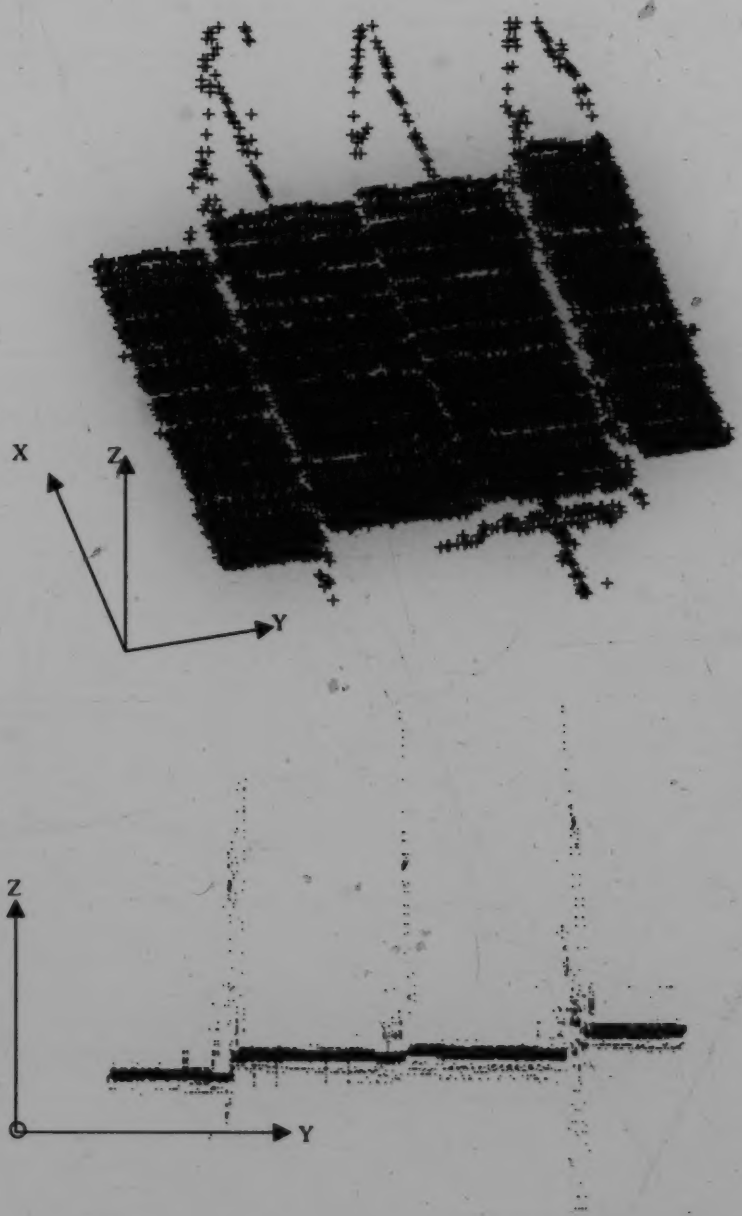


Figure 5.9: 3D view of the steps

Parameter	Known	Calculated	Error
Height	305mm	301mm	4mm
Width	305mm	303mm	2mm
Step1	10 mm	10.4 mm	0.4mm
Step1-Step2	3 mm	3.1 mm	0.1mm
Step2-Step3	10 mm	10.2 mm	0.2mm

Table 5.4: Results obtained for step object.

5.2 Analysis of Results

The results of the tests on the curved object indicate that the maximum difference between the calculated co-ordinates and the actual coordinates is 5mm and the measurements are precise to 1mm over three tests. For the triangular surface the maximum error is 5mm. However for the step object the error is 0.4mm when the size of the step is 1cm. The results obtained indicate that the feature extraction process is very precise and accurate for the curved object and for the ramp like surface. However for the step like object there are a number of points where the features are not extracted correctly especially near the steps. In figure 5.8(a) there are some points where the lines have noisy peaks. Figure 5.8(b) shows that it has straight lines and has no noise. This can be attributed to the fact that image captured from the top camera is noisy, or step like surface has caused the lighting to be distorted for the top camera. It can be observed that the points close to the steps are more prone to false feature extraction than the points at the uniform flat surface. These erroneous points have a cumulative effect on the matching stages and the triangulation stage. It can be observed from figure 5.9 that the points that were falsely extracted do not appear on the surface but are at a certain height away from the surface. This is due to the fact that these points were not matched correctly and therefore their 3-D coordinates, are incorrect. It is also observed that the other two surfaces have a few points, where the feature extraction has failed. However these points account for less than 0.3% of the total number of points (4,000 points). For the step like object the entire column along the step has erroneous results. The number of points with incorrect 3D points are 210 which is less than 2% of the total points (12,000). The number of steps increases the number of points with incorrect 3-D points due to erroneous feature extraction close to the steps.

5.3 Recovery of 3-D Surface Information from the Back Images

After the tests were conducted on known models, further testing was performed on 5 subjects (3 male subjects and 2 female subjects). The subjects selected for the testing had a range of height, weight, and build. Figure 5.10, shows the stereo images captured from a male subject. This subject has a normal trunk surface. The results of feature extraction and the 3-D surface reconstruction are shown in figures 5.11 and 5.12 respectively. The processing time was 11 minutes. (This is calculated based on the performance on a HP PA-RISC (Visualize C160) powered machine running Matlab 5.1) The region of interest was of size 400 x 300 pixels. For this subject the measurement was repeated three times. For every trial, images were captured with the patient in normal standing position and had to hold breath for less than 20 seconds to avoid any movement artifacts. Each trial was considered as a new measurement and the subject was repositioned for every trial. The aim of this experiment was to understand the effect of patient posture during measurement and how the measurement varied from one trial to another. As there were no specific markers placed on the subject during these measurements points of maximum depth change was taken as a parameter for comparison from one run to another. The shoulder blade region (Pt. A) had the maximum height and the other point chosen was close to the lumbar-spine region (Pt. B). The results of the three trials are shown in Table 5.5.

Parameter	Trial 1	Trial 2	Trial 3
Height Difference between Pt. A and Pt B.	37mm	40 mm	42mm

Table 5.5: Results for three trials conducted on male subject

However point A and point B were not accurately determined, an approximate location was chosen for each measurement. The results show that there is a maximum deviation of 5 mm from one measurement to another and this can be due to the different pose of the patient from one trial to another. Further testing is required by placing markers on the surface of the patient so that the effects of posture on the measurement from one trial to another can be studied. This would enable clinicians to identify if the

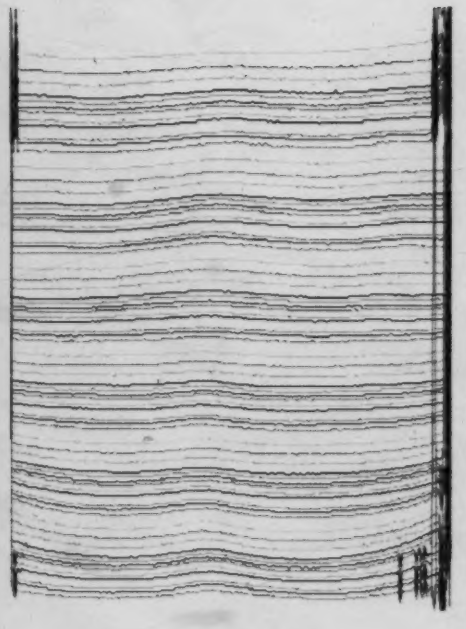
measurement taken over a period of time is different due to actual changes in the back shape or due to difference in posture of the patient from one clinical trial to another. Figure 5.13 shows the stereo images of a female subject with mild scoliosis. The results of feature extraction process and the three-dimensional surface reconstruction are shown in figure 5.14 and 5.15 respectively. As the actual coordinates on the surface of the patient of the patient was not available only a visual comparison was made from the surface obtained about the performance of the system. However it can be observed from the results of feature extraction (figure 5.14) that there are no false extractions. As there are no sudden changes in depth on the surface of the back there are no false extraction of features. The region of interest was of size 300 x 230 pixels and the processing time was 9 minutes. Stereo images of another male subject are shown in figure 5.16 and the results of the 3-D surface reconstruction are shown in figure 5.17. The region of interest for this subject was 400 x 280 and the processing time was 11 minutes. Figure 5.18 shows the top camera image of the 3rd male subject and it can observed that the surface area for this subject is relatively higher than the other subjects and the surface is predominantly flat. The region of interest for this subject was 430 x 370 and the surface reconstruction of the back is shown in figure 5.19. It can be observed that the surface closely resembles the surface of the subject (by visual inspection). The processing time for this subject was 12 minutes.

From the tests conducted on the subjects it was found that posture affects the measurement from one trial to another. Further tests are required to accurately determine this effect on the measurements. The processing time varies from 9 minutes to 12 minutes depending on the surface area (or region of interest). The clinically acceptable time is the duration of the patient's visit to the clinic, which is about 10-15 minutes. The processing time can be reduced. This is discussed in the following chapter. The results are compared by visual inspection and some measure of the actual co-ordinates are required by placing markers on the subjects during testing for determining the precision of the system. The results show that feature extraction process does not have any noise and this indicates the system's ability to extract features from the surface of back without errors. The tests performed show the feasibility of the image acquisition system in reconstructing trunk surfaces. Further investigation is required on subjects with severe scoliosis.

(a) Top Camera

(b) Bottom Camera

Figure 5.10: Stereo images of a male subject



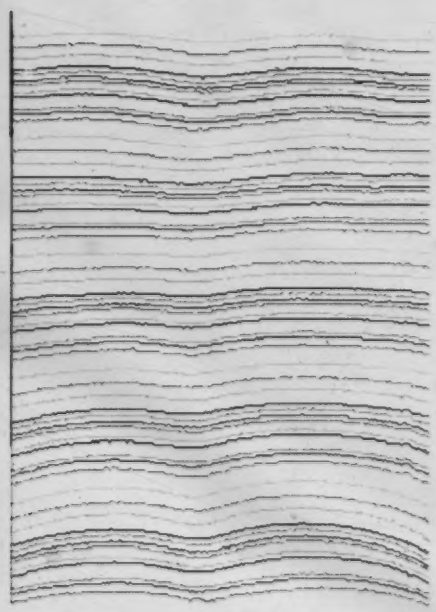


Figure 5.11: Results after feature extraction

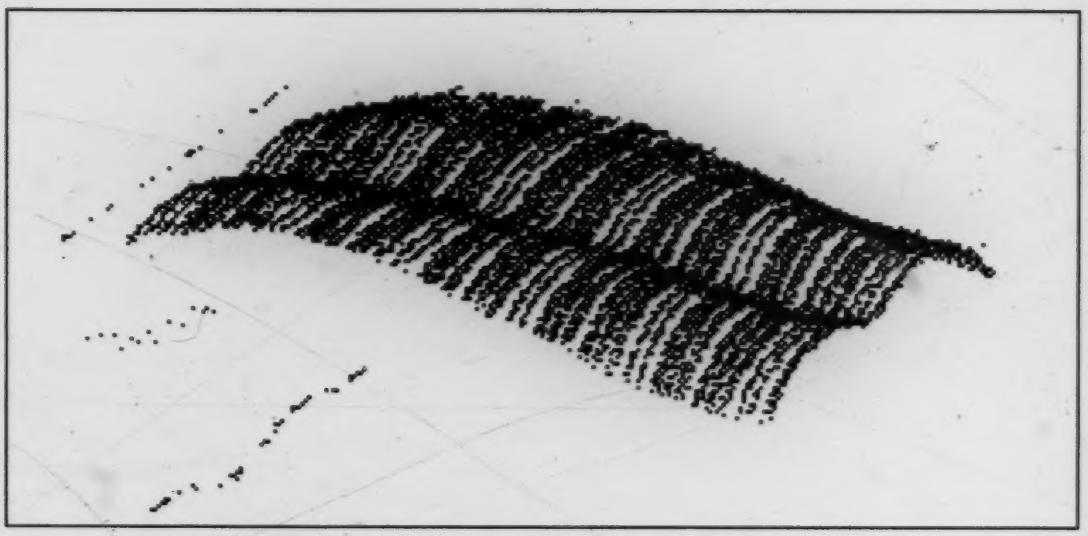
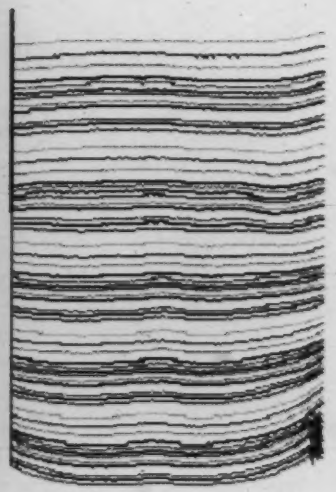


Figure 5.12: 3-D view of the reconstructed surface of the back

(a) Top Camera

(b) Bottom Camera

Figure 5.13: Stereo Images of female subject



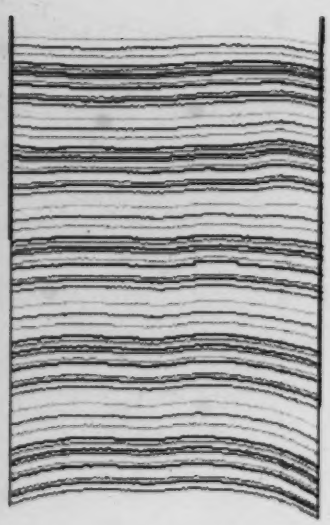


Figure 5.14: Results after feature extraction

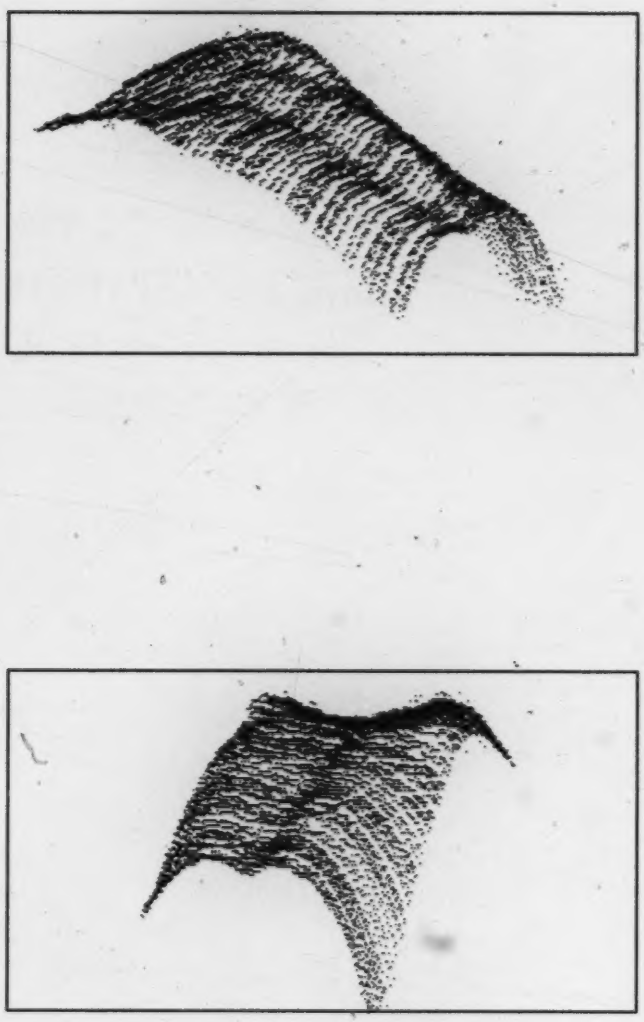


Figure 5.15: 3D view of the surface

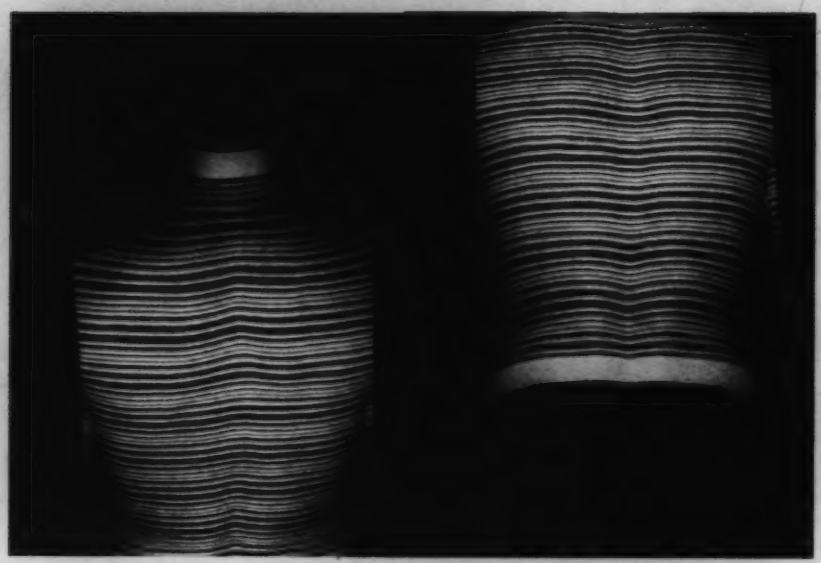


Figure 5.16: Stereo images of 2nd male subject

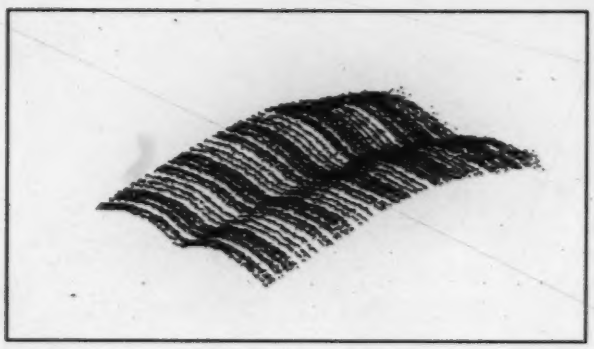


Figure 5.17: 3D view of the reconstructed surface

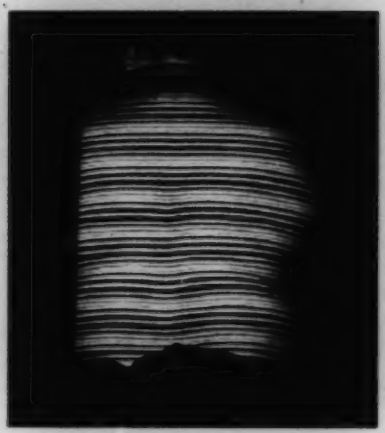


Figure 5.18: The top camera image of the 3rd male subject with wider surface area than the other subjects.

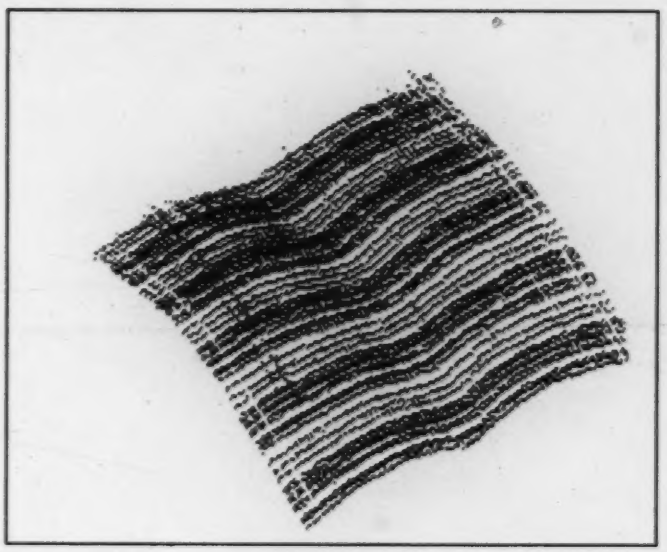


Figure 5.19 : 3-D view of the reconstructed surface

5.4 Discussion

The tests conducted on the test objects have shown that the system is capable of making precise measurements for a number of objects of different shapes with varied changes in depth and sizes. From the results obtained, the following has been observed:

System Performance:

Resolution: The line density after feature extraction is 4mm. The horizontal resolution is 1mm and the vertical resolution is 4mm.

Accuracy: The accuracy was measured with respect to the known test objects and further investigation is required to evaluate the accuracy of the system on the surface of the back. The tests have indicated that the system has higher accuracy with objects that have a curved surface or a uniform depth change than with objects which have sudden depth change like a step. The error with the object that has a step change of 1cm is 0.4mm.

Processing time: The time taken from image capture to 3-D surface reconstruction is about 12 minutes for a typical surface of the back. This is calculated based on the performance on a HP PA-RISC (Visualize C160) powered machine running Matlab 5.1. This time is dependent on the size of the region of interest that is extracted of the captured image. The correlation process is computationally very expensive and consumes about 78% of the time, as two sets of correlation are performed along each column of the image. The image rectification and image matching techniques account for the other 20%. Further studies are required to evaluate the relationship between the image size and processing time. In this technique, it has been observed that the time is proportional to the height of the image as the correlation is done columnwise, Therefore the time consumed must be calculated per column of the image with a fixed number of rows.

Level of Automation: The system is automatic from the point of image capture to 3D surface reconstruction. No user input is required. The two procedures that require manual intervention are image capture and camera calibration.

Limitations:

From the results obtained it is observed that the vertical resolution of 4mm can be achieved and the maximum error in the measurements is 5mm. The error in measurements can be attributed to a number of factors. Some of the significant factors are

Calibration:

Camera calibration is the heart of this measuring system. The results of camera calibration are used directly for image rectification as well as 3D-point calculation by triangulation. The variation in calibration data can affect the results of the measurement significantly. In the present technique the calibration is done by manually digitizing the points of the control frame in the two images. Trials have shown that this method of digitizing can yield different results for the same camera set up. It has also been observed that this inaccuracy can affect the calculation of the absolute location of X, Y, Z but does not significantly affect the relative locations of a point with respect to another. This can be due to the fact that results affect both the cameras and may be cancelled out during the matching phase. However, on the other hand if the digitizing affects the two cameras differently, then the errors add up and lead to false measurements. Therefore calibration has to be carried out with precision for accurate measurements.

Camera resolution and frame grabber performance:

As discussed in Chapter 3, the present camera resolution of 800 x 500 is just suitable for the application, however the frame grabber has a resolution of 640 x 480. Therefore the capabilities of the cameras are not fully utilized. This process of cropping down the captured images to the frame-grabber's resolution becomes another source for error in the measurement. As discussed in Chapter 4, 4mm on the object surface corresponds to 3 pixels on the captured image, therefore a ± 1 pixel jitter can have a $\pm$ 1.33mm difference in measurement.

Another important source of error is in the digitizing of the captured image by the frame grabber. The captured image has pixels with intensity level ranging from 0-255, although the projected pattern has only three distinct levels. The projector lighting can

also add a certain amount of noise to the image, but the 3 levels are scattered among 256 levels, which makes it difficult to identify the third level gray distinctly.

Effects of Projector:

The projector plays a significant role in any structured light system. The projection of a known pattern of light makes the process of correspondence matching between stereo images easier. However, if the projector distorts the pattern, then it can become a source for erroneous measurement and defeats the purpose of projecting a known pattern of light. The distortion in image projection can propagate through the entire reconstruction process and affect the accuracy of the system.

Errors introduced by Image Processing:

Although the images can be processed to filter out the noisy image and extract useful features, any kind of processing of the pixels can add to the errors as there is no way of perfectly removing only the noisy pixels. However these errors are very small in comparison to the error introduced due to calibration or image capture.

Feature extraction: The feature extraction in this approach relies on correlation techniques. The correlation technique depends on relative intensity between adjacent pixels. Improper lighting or noise in image capture can result in erroneous correlation. However it has been observed that with suitable lighting the correlation results are accurate and a worst case error of 2 pixels can be expected.

Stereo Correspondence: In this technique, the stereo matching is completely dependent upon the feature extraction process and the epipolar constraint. Any error in image rectification can affect the application of the epipolar constraint. The image rectification in turn depends on accurate calibration for the camera parameters. Therefore errors in stereo correspondence depend on accurate calibration and feature extraction.

Triangulation: The computation of the 3D points by triangulation is another process, which requires the camera calibration data. The triangulation process by itself is an

accurate process and yields precise results over a number of runs. However any error in camera calibration can carry over to the error in triangulation.

From the performance of this system it is evident that with the approach used in this research it is possible to solve the stereo matching problem to some extent. The three level coding scheme used in this research along with the feature extraction process is an improvement over the binary (line) coding scheme used before [27]. The problem of false feature extraction is reduced and the resolution of this system is higher than the 13mm vertical resolution achieved earlier. By eliminating errors in the feature extraction process the accuracy of the system has been improved. From the literature review conducted it is observed that researchers have used different approaches to solve the stereo matching problem [9], [26], [32]. Although considerable progress has been made in three-dimensional computer vision using range data there remain some fundamental problems, both analytical as well as computational. The approach used in this research can also be applied to machine vision problems in general. However the operating environment in a clinic is significantly different from the conditions on an assembly line for automatic fault analysis. This system has been designed considering the requirements of a clinician. The results also show that the approach used in this research can be used to solve the stereo matching problem in similar systems consisting of low-resolution black and white cameras.

6 Conclusion and Recommendations for future work

6.1 Conclusion

The objective of this research was to develop a coding scheme that can be used to improve the pattern extraction and matching of structured light projected on the surface of the human trunk. A three level coding scheme was developed which has a specific sequence of white, gray and black lines. Image processing software was implemented to extract the features from the projected light and reconstruct the 3-D surface of the human trunk. Validation tests were performed on objects of known dimensions and on human subjects and the performance of the system was evaluated. The task required for the development of this system was divided into 5 different modules.

- Stereo Camera calibration
- Development of a coding scheme
- Feature extraction
- Correspondence matching
- Computation of 3-D points and surface representation

For each of these tasks, a well-defined problem was formulated, literature survey was conducted and a solution was proposed and implemented. Validation tests were performed after the implementation of each stage. From the results obtained from the testing of the system it can be concluded that this approach can be developed into a reliable system for measuring the degree of scoliosis. However further testing and improvements are required in some of the above areas before it can be made available as a useful clinical tool. The areas that require further work are discussed in the following section.

6.2 Recommendations for Future Work

There are two directions of future work.

1. To use the current coding scheme and develop the system into a useful clinical tool.

The areas that require further work in this direction are:

Calibration:

The calibration procedure involves manual digitizing of the points on the control frame. The digitized points vary from one run to another with the same camera set up. This results in different calibration data for the same camera parameters. This can be avoided by automating the digitizing process. During the calibration trials conducted for this research, small differences in camera parameters significantly affected the computation of the 3-D points. Further study has to be conducted to quantify these variations and determine the exact relationship between the different camera parameters and their effect on image rectification and 3-D point computation.

Processing time:

As discussed in section 5.3 the computation time varies from 10-12 minutes for the surface reconstruction. 80% of this time is consumed by correlation. Implementing frequency domain correlation can reduce the time for correlation. To take advantage of the frequency domain correlation the template size has to be increased. One technique for increasing the template size would be by increasing the length of the projected code by having more combinations and thereby reducing the spatial frequency of the code.

The region of interest can also be determined more accurately by improving the boundary detection technique. Presently a rectangular region of interest has been chosen. By reducing the region of interest to exactly the required features, computational time can be reduced to some extent.

Image Processing:

The accuracy of the feature extraction process can be improved by using filters to remove the noise from the captured image. The filtering technique would depend on the lighting conditions since the projector lighting may change from one run to

another. Therefore the lighting conditions of the projector have to be fixed by a suitable technique before applying any filter. Another factor to be considered while applying the filter is to avoid distorting the image further. As the extracted features from the image signify the shape of the surface, the filter should not affect the extracted features in any fashion. This is a challenging problem and requires some constraints to be applied to the filter.

Surface Representation:

The 3-D surface representation has to be made user friendly by creating software that would give depth information of any point as the user moves the mouse on the 3-D surface. The surface representation of the computed 3-D points is not within the scope of this research and further work is required in this area.

2. Further research is required to improve the coding scheme by using multilevel or colour-coding schemes and thereby improve the performance of the system.

Techniques to improve the coding scheme:

As discussed in Chapter 3, the coding scheme selected for this research is limited by the camera-projector system. The three level coding scheme was found to be the most appropriate scheme for the available set-up. With high-resolution black and white cameras it should be possible to reduce the width of the lines from 4 pixels to 2 or less. Reduced line width increases the resolution of the system from the currently achieved 4mm to 2 mm. The feature extraction process is carried out by correlation, which is very sensitive to lighting conditions. The use of high-definition LCD projector would greatly improve the lighting and yield higher correlation coefficients and thereby increase the accuracy of the feature extraction process. Digital cameras can be used to decrease the $\pm$ 1 pixel jitter and the integrating effects of the frame-grabber and thereby the correlation results improved. If the captured image has no distorting effects due to the camera and the projecting device, the correlation process can be replaced by simpler feature extraction techniques. This can also reduce the computation time significantly. With higher resolution projecting devices and high-resolution cameras it is also possible to use

complex high-density codes that were discussed in section 2.3. However it is necessary to determine the desired accuracy from a surface topography measurement system to assess scoliosis.

Further research has to be conducted to study the feasibility of using colour-coded patterns. The use of colour codes would result in tremendous increase in the number of possible combinations and thereby much higher resolutions can be achieved. With the availability of low-cost high-resolution colour digital cameras, the benefit/cost ratio is also greatly increased and should be considered as serious replacement to the present system.

References

[1] K.Bhalla, "A stereo vision System for the intraoperative Monitoring of Scoliosis Correction," *M.Sc Thesis, Dept. of Electrical Engg.*, University of Alberta, 1995.

[2] K.L.Boyer and A.C.Kak, "Colour-Encoded structured light for rapid range sensing", *IEEE Trans. PAMI*, Vol. PAMI-9, no. 1, pp. 14-28, Jan. 1987.

[3] A.Busboom, R.J.Schalkoff, "Active stereo vision and direct surface parameter estimation: curve to curve image plane mappings," *IEE Proceedings-Vision Image Signal processing*, vol.143, no. 2, pp. 109-117, April 1996.

[4] J.F.Cardenas-Garcia, H.G.Yao, S.Zheng, " 3D Reconstruction of objects using stereo imaging," *Optics and Lasers in Engineering*, vol. 22, no. 3, pp. 193-213, 1995.

[5] Chu-Song Chen, Yi-Ping Hung, Chiann-Chu Chiang, Ja-Ling Wu, " Range data acquisition using colour structured lighting and stereo vision," *Image and Vision Computing*, Vol. 15, pp. 445-456, June 1997.

[6] Datong Chen, Wen Gao , Xilin Chen, "A new approach of recovering 3-D shape from structured lighting," Proc. Of ICSP' 96, pp. 839-842, 1996.

[7] Tsorng-lin Chia , Zen Chen, "Locating a partial cylindrical surface from a structured light image," *Proc. Natl. Sci. Counc. ROC(A)*, vol. 18, no.1, pp. 85-92, 1994.

[8] W.Cho. "Transformation of the pixel system to the vertical position and resampling," *Technical notes in photogrammetry 3*, Ohio State University, Dept. of Geodetic Science and Surveying, Columbus Ohio, 1989.

[9] S.L.Chou and W.H.Tsai, "Line segment matching for 3-D computer vision using a new iteration scheme," *Proc. Int. Computer Symposium*, Hsinchu, Taiwan, R.O.C pp. 414-421, 1990.

[10] Ahmed M. Darwish, " 3D from Focus and Light stripes," *SPIE Sensors and Control for Automation*, vol. 2247, pp. 194-201, 1994.

[11] I.Dobelle, "Artificial Vision for the Blind: The Summit May Be Closer Than You Think," *Transactions of the American Society for Artificial Internal Organs*, vol. 40, no.4, December, 1994.

[12] G.L.Foresti, "Recognition of 3D moving objects for autonomous vehicle driving," *28th Int. Symposium on Automobile Tech. And Automation, Proc. of the Dedicated Conf. On Robotics, Motion and Machine Vision in the Automotive Industries*, pp. 83-90, 1995.

[13] H.Gartner, P.Lehle, Tizini.H.J and C.Voland, "Structured light measurement by double scan technique," *Proc. SPIE*, vol. 2784, pp. 21-30, 1996.

[14] H.Gartner, P.Lehle, Tizini.H.J, "New, highly efficient, binary codes for structured light methods", *Proc. SPIE*, vol. 2599, pp. 4, 1996.

[15] Jorg Gerber, Peter Kumnstedt.Richard Kowarschik, Gunther Notni, Wolfgang Schreiber, "3D-coordinate measuring system with structured light," *SPIE Photomechanics*, vol. 2342, pp. 41-49, 1994.

[16] R.C.Gonzalez and R.E.Woods, "*Digital Image processing*," Addision-Wesley Publishing Company, 1992.

[17] P.M.Griffin, Lakshmi S. Naraimhan and Soung R.Yee, "Generation of uniquely encoded light patterns for range data acquisition," *Pattern recognition* vol. 25, no. 6, pp. 609-616, 1992.

[18] W.E.L.Grimson, "*From images to surfaces: a Computational study of Human Early visual system*," MIT press, Cambridge, Massachusetts, 1981.

[19] Robert M.Haralik, Linda G. Shapiro, "*Computer and Robot Vision Vol I*," Addison-Wesley Publishing Company, 1992.

[20] Tomita Hiroaki, Shigeru Yoda, Kosuke Sato and Kunihiro Chihara, "Range finding with simultaneous projection of modulated patterns," *Proc. SICE*, pp. 1449-1452, 1995.

[21] Gongxhu Hu , George Stockman, "3-D surface solution using structured light and constraint propagation," *IEEE Trans. PAMI*, vol. PAMI-2, no.4 , pp. 390-402, April 1989.

[22] Seiji Inokuchi, Kosuke Sato and Fumio Matsuda, " Range-Imaging system for 3-D object recognition," *Int. Conf. Pattern Recognition*, pp. 806-808, 1984.

[23] R.A.Jarvis, "A perspective on range finding techniques for computer vision," *IEEE Trans. PAMI*, vol. PAMI-5 no.2 pp. 122-139, Mar.1985.

[24] T.J.Keating, Wolf, P.R., and Scarpace, F.L. "An Improved Method of Digital Image correlation," *Photogrammetric Engineering and Remote sensing*, Vol. 41, No.8, pp. 993-1002, 1975.

[25] Philippe Lavoie, Dan Ionescu, Emil M. Petriu, " 3-D Model recovery from 2-D images using structured light," *IEEE Instrumentation and Measurement Technology*, pp. 377-382, 1996.

[26] Shing-Huan Lee , Jin -Jang Leou, "A dynamic Programming approach to line segment matching in stereo vision," *Pattern recognition*, vol. 27, No. 8, pp. 961-986, 1994.

[27] R.Liu, "Three Dimensional Reconstruction of the Trunk Surface Using Structured Light," *M.Sc Thesis, Dept. of Electrical Engg.*, University of Alberta, 1995.

[28] D.Marr and T.Poggio, "A theory of human stereo vision," *Proc. R. Soc. London*, vol. B204, pp. 301-328, 1977.

[29] Ding Mingyune, Friedrich M.Wahl, "Using space continuity and orientation constraints for range data acquisition," *Pattern Recognition*, vol. 27, no. 8, pp. 987-1004, 1994.

[30] Jacqueline Moigne and Allen M.Waxman, " Projected light grids for short range navigation of autonomous robots," *Proc. 7th, Int. Conf. Pattern Recognition*, pp. 203-206, July 1984.

[31] Hiroyoshi Morita, Kazuyasu Yajima and Shojiro Sakata, "Reconstruction of surfaces of 3-D objects by M-array pattern projection method," *IEEE International Conference on Computer Vision*, pp. 468-473, 1988.

[32] Y.Ohta and T.Kanade, "Stereo by intra- and inter scanline search using dynamic programming," *IEEE Trans. PAMI*, vol. PAMI-7 no.2, pp. 139-154, 1985.

[33] J.C.Perrin, and Thomas, A, "Electronic processing of moiré fringes: Application to moiré topography and comparison with photogrammetry," *Appl. Opt.* vol.18, no.4, pp. 563-574, 1979.

[34] A.E.Peterson, N.G.Durdle, V.J.Raso and D.L.Hill, " Calibration of Video cameras for Scoliosis mapping," *Geomatica*, Vol 47, pp. 29-38, 1993.

[35] J.L.Posdamer and M.D.Altschuler, "Surface Measurement by space encoded projected Beam Systems," *Computer Graphics and Image Processing*, vol.18, pp. 1-17, 1982.

[36] M.Potmesil and H.Freeman, "Curved surface representation utilizing data extracted from multiple photographic images," *Workshop on the representation of Three-Dimensional objects*, Philadelphia, PA, pp. H1-H26, May 1979.

[37] W.H.Press, W.T.Vetterling, S.A.Teukolsky, and B.P.Flannery, "*Numerical Recipes in C: The Art of Scientific Computing*," Cambridge University Press, second edition, 1992.

[38] Giovanna Sansoni, Stefano Corini, Sara Lazzari, Roberto Rodella, and Franco Docchio, " *3-D Imaging of surfaces for Industrial application : integration of structured light projection , Gray code projection and projector-camera calibration for improved performance*," SPIE, vol. 2661, pp. 88-96, 1996.

[39] E.Schubert, Rath.H, Klicker.J, "Fast 3d object recognition using a combination of colour-coded phase-shift principle and colour-coded triangulation," *SPIE Vol.2247 Sensors and Control for Automation*, pp. 202-213, 1994.

[40] G.Stockman and Gongzhu Hu, "Sensing 3-D surface patches using a projected grid," *IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Miami Beach, FL, pp. 602-607, 1986.

[41] P.Vuylsteke. , Oosterlinck.A, "Range Image Acquistion with a single Binary-Encoded light pattern," *IEEE Trans. PAMI*, Vol 12, No. 2, pp. 148-163, February 1990.

[42] Y.F.Wang and J.K.Aggarwal J.K, " An overview of geometric modelling using active sensing," *IEEE Control Systems Magazine*, pp. 5-12, June 1988.

[43] Y.F.Wang. Y.F.Mitiche.A, Aggarwal, J.K, "Computation of surface orientation and structure of objects using grid coding," *IEEE Trans. PAMI* , vol. PAMI-9, no.1, pp.129-137. Jan 1987.

[44] P.M.Will, Pennington.K.S, "Grid Coding: A novel Technique for Image Processing," *Proceedings of the IEEE,* Vol. 60, No.6, pp. 669-680, June 1972.

[45] P.R.Wolf. "*Elements of Photogrammetry*," McGraw-Hill Inc, New York, U.S.A, 2nd Edition, 1983.

[46] C.Wust and David .W.Capson, "Surface Profile measurement using Colour Fringe projection," *Machine Vision and Applications* , vol. 4, no.3, pp. 193-203, 1991.

[47] H.S.Yang, Boyer.K.L. Kak A.C. "Range data extraction and interpretation by structured light," *Proc. Int. Join. Conf. Artif. Intell.*, pp. 199-205,1984.

Appendix

The appendix briefly describes the system specifications in terms of the software developed and platform used. The description of each module has also been provided.

Software Information:

Programming language: ANSI C, and MATLAB 5.1

Operating system: UNIX

Platform : HP-UX, (Note: MATLAB Modules can also be run on a PC)

Software available with: Dr. N.G.Durdle, Dept. of Electrical Engineering, University of Alberta, Edmonton, Alberta, Canada T6G 2G7.

Module Descriptions:

A high level description of the individual modules in terms of their purpose, input and output are provided here. All the modules are independent. Further details about each module, their individual functions, data structures and variables are provided in the introductory comments of each code and README files written for future programmers. The flow of the software modules are shown in figure A.1.

Calibration Software: This calibrates the two stereo cameras and the parameters of the camera are stored in file cam.dat. The software details for calibration are available with Prof. A.E.Peterson, Dept. of Civil and Enviro., Engg., University of Alberta, Edmonton.

Directory name: ~olson/calibration

Executable names: camera_calib, ndlt, v2stereo

Tiff-data conversion: The captured tiff stereo images are converted to data files using this module. The input image files for this module should be in pgm format.

Directory name: slides

Executable name: mainwo

Image format supported: pgm format

Input: te.pgm, The captured images in tiff format have to be converted to pgm format for the program mainwo.c

Output: "filename".dat

Image Rectification Module: This is the module that rectifies the two images that are captured.

Program: rectfy.m

Input: The inputs to this module are the camera calibration date file cam.dat and the two stereo image data. The tiff data is converted to data file by using the program mainwo.c.

Output: Two rectified image data O1 and O2.

Feature Extraction module: This module is responsible for extracting the lines from the rectified images. The region of interest is determined by the regofi.m function and this module calls the correlation functions to perform template matching.

Program: featextr.m

Input: O1 and O2

Output: Locations of extracted features for each column are stored for each image as fext1 and fext2

Correspondence matching: This module matches the features extracted from each image and generates a file of the locations of matched pairs.

Program: prespaco.m

Input: fext1 and fext2

Output: try.txt

Triangulation: This module computes the 3-D coordinates of the matched points.

Program: spaco (Fortran executable).

Input: camera calibration data, cam.dat and matched points try.txt.

Output: SAVE.TMP

3-D surface representation: This module displays the 3-D points computed. Any angle of view can be selected.

Program: surfrep.m

Input : SAVE.TMP

Output: No files are generated. This plots the 3-D points on screen.

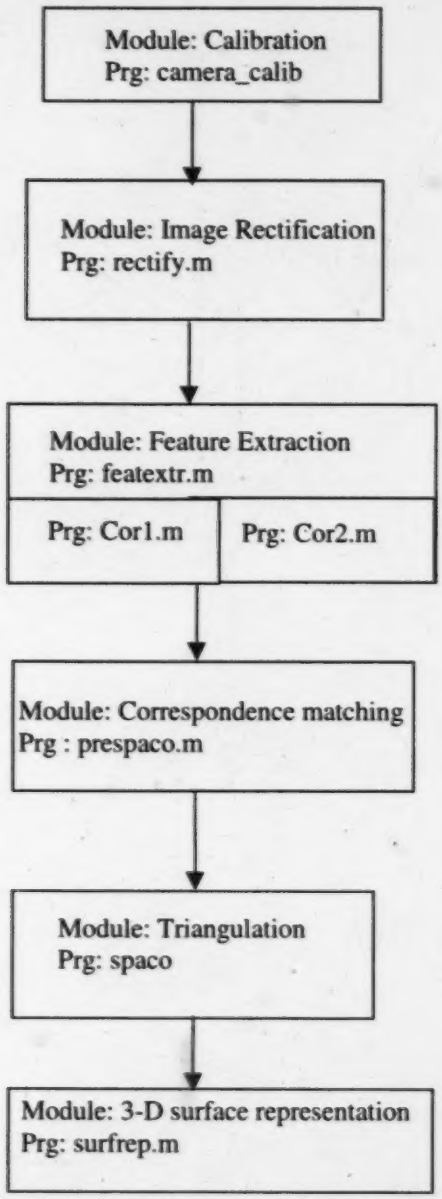


Figure A.1: Flow of the software modules.